Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 24 (2022), Issue 7 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Editorial

Reviews

Viewpoints

Original Papers

Letters to the Editor

Corrigenda and Addenda

Short Paper

Research Letter

Editorial

Brilliant Ideas Can Come in All Sizes: Research Letters

Rita Kukafka[1*], DrPH, MA; Tiffany I Leung[2*], MD, MPH; Gunther Eysenbach[2,3*], MD, MPH

[1]Columbia University Irving Medical Center, New York, NY, United States
[2]JMIR Publications, Inc, Toronto, ON, Canada
[3]University of Victoria, Victoria, BC, Canada
[*]all authors contributed equally

Corresponding Author:
Tiffany I Leung, MD, MPH
JMIR Publications, Inc
130 Queens Quay East, Unit 1100
Toronto, ON, M5A 0P6
Canada
Phone: 1 416 583 2040
Email: tiffany.leung@jmir.org

Abstract

The *Journal of Medical Internet Research* is pleased to offer "Research Letter" as a new article type. Research Letters are similar to original and short paper types in that they report the original results of studies in a peer-reviewed, structured scientific communication. The Research Letter article type is optimal for presenting new, early, or sometimes preliminary research findings, including interesting observations from ongoing research with significant implications that justify concise and rapid communication.

(J Med Internet Res 2022;24(7):e41046) doi:10.2196/41046

KEYWORDS

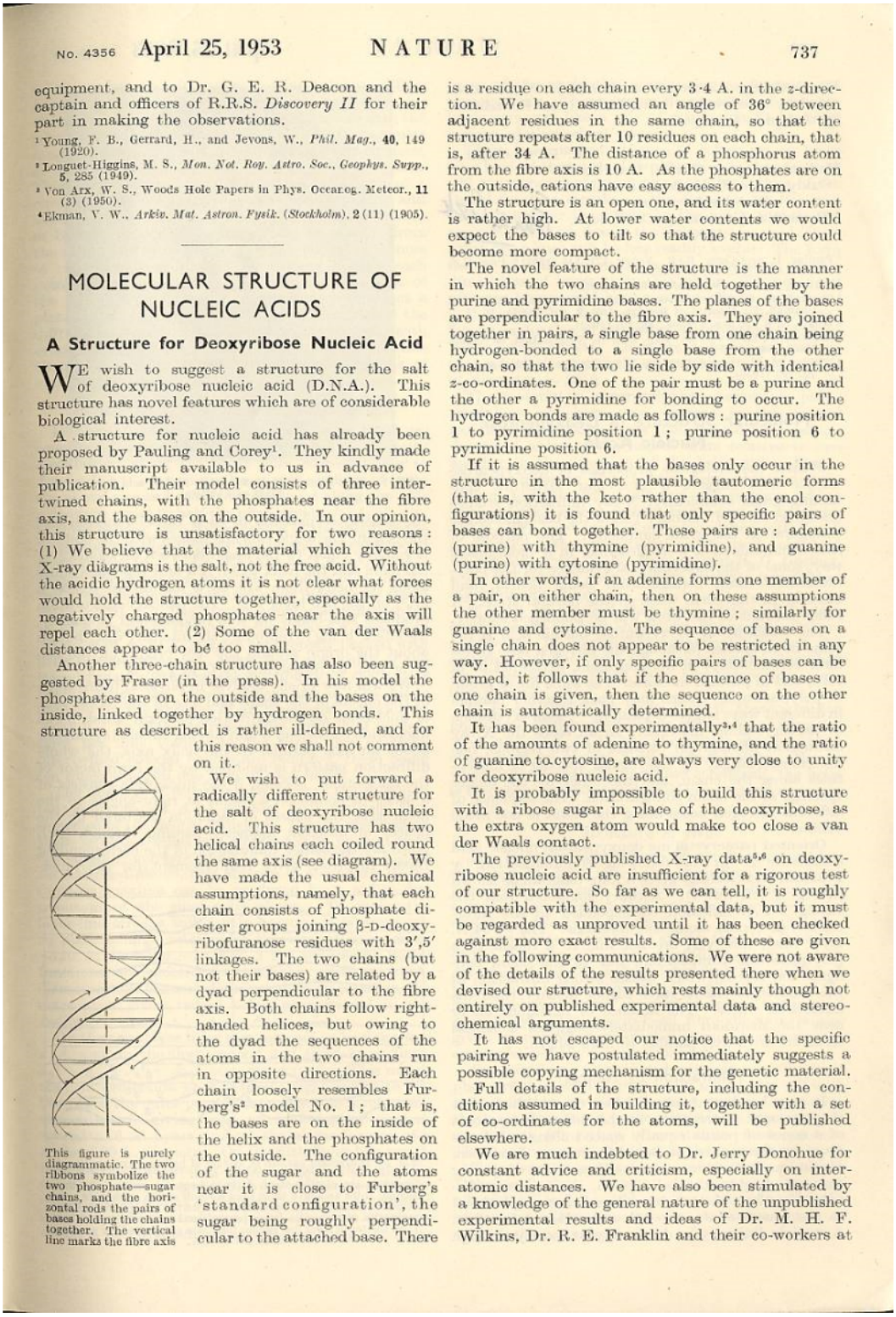
open science; open access publishing; publishing; scholarly publishing; scientific publishing; research; scientific research; research letter

Did you know that Albert Einstein published his famous $E=mc^2$ equation on mass-energy equivalence in roughly 2 pages [1]? Or that the original and preliminary communication suggesting the double-helix structure of DNA by Watson and Crick (Figure 1 [2,3]) is also only a little more than 1 page in length? If winning a Nobel prize is evidence of brilliance, then one may conclude that the length of a manuscript is not commensurate with its value.

Because less is sometimes more, the *Journal of Medical Internet Research* is now pleased to offer "Research Letter" as a new article type. Research Letters in the *Journal of Medical Internet Research* are similar to original and short paper types in that they report the original results of studies in a peer-reviewed, structured scientific communication. The Research Letter article type is optimal for presenting new, early, or sometimes preliminary research findings, including interesting observations from ongoing research with significant implications that justify concise and rapid communication.

The *Journal of Medical Internet Research* is publishing Research Letters for several reasons. First, the Research Letter is an optimal medium for quickly communicating transformative work, offering authors an opportunity to submit their focused research work for potentially more rapid peer review and publication processes simply by the nature of the communication. Second, larger and more extensive research on contemporary issues might also produce focused findings that may be incidental to the primary aims, yet still be valuable to report. One interesting key result can be displayed in 1 or 2 tables or figures. Additionally, students and early career researchers are encouraged to submit Research Letters as a pathway for reporting their impactful, targeted research projects; this may offer a stepping stone for these researchers as they publish work that contributes to the field and to their scientific growth and professional advancement. For readers, who often include busy scientists and professionals, Research Letters can offer new ideas or approaches in a brief and quickly digestible, yet robust and high-quality, manner. Taking experiences from other high-impact journals, Research Letters are often highly cited.

Figure 1. Archived scan of "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid," published on April 25, 1953, by Watson and Crick [2]. Source: Linus Pauling and the Race for DNA [3].



No. 4356 April 25, 1953 NATURE 737

equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.R.S. *Discovery II* for their part in making the observations.

1 Young, F. B., Gerrard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1920).
2 Longuet-Higgins, M. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, **5**, 285 (1949).
3 Von Arx, W. S., Woods Hole Papers in Phys. Oceanog. Meteor., **11** (3) (1950).
4 Ekman, V. W., *Arkiv. Mat. Astron. Fysik. (Stockholm)*, **2** (11) (1905).

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

This figure is purely diagrammatic. The two ribbons symbolize the two phosphate—sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining β-D-deoxyribofuranose residues with 3′,5′ linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3·4 A. in the z-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical z-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[5,6] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on interatomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at

Research Letters should still present original work that has not been previously published. Work presented at a conference that has not been previously published in proceedings can be submitted as a Research Letter. However, tables or figures from previously published or submitted papers would not be considered in a Research Letter. Authors can refer to article type information on the format of a Research Letter in JMIR Publication's Knowledge Base [4]. In this issue of the *Journal of Medical Internet Research*, the journal has published its first example [5], with additional Research Letters currently in review.

We encourage authors to consider submitting their Research Letters to the *Journal of Medical Internet Research*. Additionally, the journal editors may suggest to authors the Research Letter article type as a more suitable format for their work. This is not intended to undersell the contribution of the

submission. Authors may not realize that the Research Letter is subject to the same rigorous peer-review process as other article types here at JMIR Publications. As we have seen from Einstein and other eminent Nobel Prize winners, brilliant ideas can be expressed succinctly.

We look forward to reviewing and publishing your Research Letters!

Conflicts of Interest

RK is the Co-Editor-in-Chief at JMIR Publications. TIL is a scientific editor at JMIR Publications. GE is founder and president of JMIR Publications.

References

1. Einstein A. The Collected Papers of Albert Einstein (1879-1955). Princeton, NJ: Princeton University Press; 1987.
2. Watson JD, Crick FHC. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. Nature 1953;171:737-738. [doi: 10.1038/171737a0]
3. Published Papers and Official Documents. Linus Pauling and the Race for DNA | Special Collections & Archives Research Center, Oregon State University Libraries. URL: http://scarc.library.oregonstate.edu/coll/pauling/dna/papers/corr68.11-reprint-19530425-01-large.html [accessed 2022-07-18]
4. JMIR Publications. What are the article types for JMIR journals? JMIR Publications Knowledge Base and Help Center. 2022. URL: https://support.jmir.org/hc/en-us/articles/115004950787-What-are-the-article-types-for-JMIR-journals- [accessed 2022-07-13]
5. Kumar A, Ren J, Ornstein KA, Gliatto PM. Using machine learning to efficiently vaccinate homebound patients against COVID-19: a real-time immunization campaign. J Med Internet Res 2022 Jul 12;24(7):e37744 [FREE Full text] [doi: 10.2196/37744] [Medline: 35679053]

Edited by A Mavragani; submitted 16.07.22; this is a non–peer-reviewed article;accepted 16.07.22; published 26.07.22.

Please cite as:
Kukafka R, Leung TI, Eysenbach G
Brilliant Ideas Can Come in All Sizes: Research Letters
J Med Internet Res 2022;24(7):e41046
URL: https://www.jmir.org/2022/7/e41046
doi:10.2196/41046
PMID:35881444

Review

The Influence of Wearables on Health Care Outcomes in Chronic Disease: Systematic Review

Graeme Mattison[1,2,3], MBChB; Oliver Canfell[1,3,4,5], PhD; Doug Forrester[1,2], PhD; Chelsea Dobbins[1,6], PhD; Daniel Smith[2], PhD; Juha Töyräs[6,7,8], PhD; Clair Sullivan[1,2,4], PhD

[1]Queensland Digital Health Research Network, Global Change Institute, The University of Queensland, Brisbane, Australia

[2]Metro North Hospitals and Health Service, Brisbane, Australia

[3]Digital Health Cooperative Research Centre, Australian Government, Sydney, Australia

[4]Centre for Health Services Research, Faculty of Medicine, The University of Queensland, Brisbane, Australia

[5]University of Queensland Business School, Faculty of Business, Economics and Law, The University of Queensland, Brisbane, Australia

[6]School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, Australia

[7]Department of Applied Physics, University of Eastern Finland, Kuopio, Finland

[8]Science Service Center, Kuopio University Hospital, Kuopio, Finland

Corresponding Author:
Graeme Mattison, MBChB
Queensland Digital Health Research Network
Global Change Institute
The University of Queensland
Cnr Butterfield St & Bowen Bridge Rd
Herston
Brisbane, 4029
Australia
Phone: 61 468933863
Email: uqgmatti@uq.edu.au

Abstract

Background: Chronic diseases contribute to high rates of disability and mortality. Patient engagement in chronic disease self-management is an essential component of chronic disease models of health care. Wearables provide patient-centered health data in real time, which can help inform self-management decision-making. Despite the perceived benefits of wearables in improving chronic disease self-management, their influence on health care outcomes remains poorly understood.

Objective: This review aimed to examine the influence of wearables on health care outcomes in individuals with chronic diseases through a systematic review of the literature.

Methods: A narrative systematic review was conducted by searching 6 databases for randomized and observational studies published between January 1, 2016, and July 1, 2021, that included the use of a wearable intervention in a chronic disease group to assess its impact on a predefined outcome measure. These outcomes were defined as any influence on the patient or clinician experience, cost-effectiveness, or health care outcomes as a result of the wearable intervention. Data from the included studies were extracted based on 6 key themes, which formed the basis for a narrative qualitative synthesis. All outcomes were mapped against each component of the Quadruple Aim of health care. The guidelines of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement were followed in this study.

Results: A total of 30 articles were included; studies reported 2446 participants (mean age: range 10.1-74.4 years), and the influence of 14 types of wearables on 18 chronic diseases was presented. The most studied chronic diseases were type 2 diabetes (4/30, 13%), Parkinson disease (3/30, 10%), and chronic lower back pain (3/30, 10%). The results were mixed when assessing the impact on a predefined primary outcome, with 50% (15/30) of studies finding a positive influence on the studied outcome and 50% (15/30) demonstrating a nil effect. There was a positive effect of 3D virtual reality systems on chronic pain in 7% (2/30) of studies that evaluated 2 distinct chronic pain syndromes. Mixed results were observed in influencing exercise capacity; weight; and biomarkers of disease, such as hemoglobin A_{1c}, in diabetes. In total, 155 outcomes were studied. Most (139/155, 89.7%) addressed the *health care outcomes* component. This included pain (11/155, 7.5%), quality of life (7/155, 4.8%), and physical

function (5/155, 3.4%). Approximately 7.7% (12/155) of outcome measures represented the *patient experience* component, with 1.3% (2/155) addressing the *clinician experience* and *cost*.

Conclusions: Given their popularity and capability, wearables may play an integral role in chronic disease management. However, further research is required to generate a strong evidence base for safe and effective implementation.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42021244562; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=244562

(J Med Internet Res 2022;24(7):e36690) doi:10.2196/36690

KEYWORDS

wearable; chronic disease; health care outcome

Introduction

Chronic diseases account for 73% of deaths worldwide [1]. The World Health Organization categorizes chronic diseases into the following four main categories: cardiovascular diseases, cancers, chronic respiratory diseases, and diabetes [2]. Half of the people with chronic disease experience disability, which is defined as a limitation that restricts daily activities and lasts for at least 6 months. Disability results in increased dependence on social services [3] and reduced quality of life [4]. Chronic disease is responsible for a significant economic burden arising from direct health care costs and lost productivity because of illness and death. An estimated 36% of all allocated health care expenditure is directed at supporting individuals with chronic diseases [5].

Involving individuals with chronic diseases in active self-management programs is recognized as an essential component of chronic disease management [6-8]. Implementing self-management programs presents numerous challenges, including limited funding, awareness, and adherence to prescribed self-management strategies [9]. A key strategic priority area in the National Strategic Framework for Chronic Conditions of Australia [10] is active engagement; that is, providing a person-centered approach that puts people at the center of their own health care and empowers them to play an informed role according to their interests and abilities. The Quadruple Aim of health care [11] provides a structured model for an approach to health care that focuses on improving the individual experience of health care delivery by improving population health, minimizing health expenditure, and maintaining the well-being of health care providers. Implementing the principle of self-management into chronic disease programs to achieve the goals of the Quadruple Aim may optimize their intended outcomes.

Patient engagement is essential to satisfy the Quadruple Aim and promote self-management in chronic disease management. Part of this engagement process involves providing patients with autonomy over their own care, including active involvement in decision-making on treatment choices and healthy lifestyle changes. Commercially available technology, including wearable devices (*wearable*) and mobile apps, can provide users with feedback on their physiological parameters, which may give them more awareness of their condition [12]. Wearables are defined as sensory devices that can be attached to clothing or worn as an accessory, which allows the tracking of health information through a multitude of onboard sensors without hindrance [13]. Initially designed for the health and fitness industry to track wellness [14], most commercially available wearables can be used to monitor key health-related metrics, including heart rate, sleep quality, energy expenditure, and step count.

An evolving area of research is the integration of wearables into the support of individuals with chronic diseases by promoting self-management strategies [15,16]. Wearables unlock the capability to perform the continuous, real-time monitoring of health status [17]. This provides a comprehensive analysis of the individual's overall health, which can be presented in a user-friendly format to patients and clinicians [18]. The ability to monitor health status remotely also strengthens integration into existing telehealth models of care, with the hypothetical capability of reducing in-person consultations between patients and clinicians [19].

Several validity studies have demonstrated early promise in the application of wearables for individuals with chronic diseases, including the prevention and treatment of cardiovascular disease [20], monitoring severity of Parkinson disease [21], and promoting adherence to exercise goals in diabetes mellitus and chronic obstructive pulmonary disease (COPD) [22]. However, implementing these devices into existing health care models will require strong empirical evidence supporting the efficacy of wearables on health care outcomes, clear implementation guidance, and research funding models [23]. Research on wearables in health care is currently in its infancy, with most studies adopting an observational research design, having small sample sizes, or focusing on healthy individuals [24]. There is minimal known, synthesized evidence for the influence of wearables on health care outcomes for individuals with chronic diseases.

To address this research gap, we conducted a systematic review addressing the question of the role of wearables in improving health care outcomes in chronic diseases. Our hypothesis is that a qualitative synthesis of the limited evidence to date may demonstrate early promise for wearables to positively influence health care outcomes, as defined by the Quadruple Aim. Our aim was to examine the influence of wearables on health care outcomes in patients with chronic diseases through a systematic review of the literature. The Quadruple Aim has been used to define health care outcomes as an internationally validated framework to guide approaches to improving health care service delivery. This research is relevant to health care researchers and

clinicians exploring the capability of wearables in health care, as well as health system managers and the wearable technology industry.

Methods

Design

A systematic review design using qualitative methods was used. We adhered to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [25]. Our PRISMA checklist is provided in Multimedia Appendix 1. This review was registered with the PROSPERO (International Prospective Register for Systematic Reviews) on April 22, 2021 (CRD42021244562).

Search Strategy

A search of studies published between January 1, 2016, and July 1, 2021, was performed using PubMed, EMBASE, Web of Science, Scopus, CINAHL, and Cochrane Central Register of Controlled Trials. Studies published before 2016 were excluded to reflect the rate of technological advancement in the research and development of wearables [26]. Our strategy was developed in consultation with a medical research librarian. A combination of Medical Subject Headings (MeSH) terms and free text keyword terms was used, including *chronic disease* (MeSH), *wearable electronic devices* (MeSH), *health care OR outcome**, and select chronic conditions such as *asthma*. Our full search strategy is available in Multimedia Appendix 2.

Eligibility Criteria

Chronic disease is defined as any health condition lasting ≥3 months, which may lead to other health complications and may be associated with functional impairment or disability [27]. A health care outcome is defined as any parameter that demonstrates an effect on the patient experience, health care outcome (such as improvement in glycemic control in diabetes), clinician experience, or cost of health care delivery.

The inclusion criteria were (1) randomized controlled trials (RCTs) and observational studies, (2) feasibility studies observing the effect of wearables on a predefined health care outcome, and (3) studies published in peer-reviewed journals in English. Studies involving adults and children were included.

The exclusion criteria were (1) pregnant patient population, (2) studies demonstrating the validity or technological feasibility of wearables, (3) studies reporting the accuracy of wearables as their primary aim, (4) books or book chapters, (5) conference abstracts, and (6) systematic reviews.

Screening

Screening of potentially eligible studies was performed in 3 steps: duplicate removal, title and abstract screening, and full-text screening. Duplicates were removed using EndNote (version 20; Clarivate) and Covidence [28]. Additional duplicates that were not removed during this process were removed manually. A total of 2 review authors (GM and DF) independently screened the titles and abstracts for inclusion using the predefined inclusion and exclusion criteria specified previously. All studies not discarded through this process were then screened via a full-text review process by 2 review authors (GM and OC), from which studies were identified for inclusion. Conflicts that could not be resolved between the 2 review authors were resolved by a third investigator (CS). Full data extraction, categorization, and labeling of papers were performed by 1 author (GM) and validated by a second author (OC).

Risk of Bias Assessment

A risk of bias assessment was conducted for all RCTs by 1 author (GM). The Cochrane Collaboration's tool for assessing the risk of bias [29] was used to assess each randomized study for bias from the randomization process, selection bias, bias because of missing outcome data, bias because of measurement of the outcome, and other biases otherwise not addressed. These are presented in Multimedia Appendix 3 [30-52]. For nonrandomized studies, the Risk of Bias in Nonrandomized Studies of Interventions tool [53] was used to assess bias because of confounding, selection bias, bias in classification of interventions, bias because of deviations from the intended interventions, bias because of missing data, bias in the measurement of outcomes, and selection of result bias. These are presented in Multimedia Appendix 4 [54-59].

Data Extraction and Synthesis

All data were extracted from the identified studies under 6 key extraction themes that were most suited to address our original research question [60] (Textbox 1). A wide range of subheadings was selected, given the high variance in disease populations and outcome measures. A narrative qualitative synthesis of the included studies was conducted. The high heterogeneity of study designs, disease groups, patient populations, and outcome measures precluded the completion of the meta-analysis. Our results are based on the disease group, with relevant findings across all studies reported within the disease group in question. All outcomes were categorized as a component of patient experience, clinician experience, health care outcomes, or cost in alignment with the Quadruple Aim. A robust assessment of the synthesis is subsequently presented through critical reflection.

Textbox 1. Data extracted from included studies (adapted from Institute of Medicine standards [60]).

General information

- Study ID, title, lead author contact details, citation, study funding sources, and country in which the study was conducted

Characteristics

- Aim, study design, start or end date, possible conflict of interest for authors, recruitment procedures used, population demographic, chronic disease studies, inclusion or exclusion criteria, and total number of participants

Intervention and setting

- Setting, intervention, cointervention (if any), control (if any), number of participants enrolled, number of participants in analysis, and number of withdrawals or exclusions lost to follow-up

Outcome data

- Unit of assessment, statistical analysis used, primary outcome (nature and measurement used), secondary outcomes (nature and measurements used), and length of follow-up

Results of study analysis

- Results with regards to primary, secondary, and exploratory outcome measures

Additional information

- Costs (if known), resources used (if known), and adverse events (if any)

Results

Classification of Wearables

The wearables used in these studies were designed to be worn either continuously or intermittently on the body. Their capabilities and commercial availabilities are summarized in Table 1. Of the 30 publications, 19 (63%) studied wearables placed on the waist (n=10, 53%) and wrist (n=9, 47%), with 3 (10%) publications studying wearables placed on multiple sites of the body. Multiple types of wearables were studied, including pedometers, smartbands, virtual and augmented reality (AR) systems, flash glucose monitoring systems, and intelligent shoe insoles.

Table 1. Wearable technology for the management of chronic diseases identified in studies within this review.

Site worn	Wearable technology	Chronic disease studied	Capability	Coexisting mobile app	Commercial availability
Arm	Inertial motion unit—shoulder and forearm placement [41]	Stroke	Movement feedback in upper limbs and estimation of angular displacement	No	No
Arm	SenseWear armband (model MF-SW) [55]	Progressive multiple sclerosis	Skin temperature measurement, limb motion detection, step count, and metabolic equivalents	No	No
Chest	Chest-worn wearable sensor [49]	Ischemic heart disease	Heart rate, respiratory rate, electrocardiogram measurement, and accelerometry	Yes	Yes
Foot	SurroSense Rx intelligent insole [44]	Diabetes mellitus with diabetic foot ulceration	Measurement of static plantar pressure	No	No
Hand	Kinesia—finger placement [50]	Parkinson disease	Feedback on finger movement	No	No
Head	EaseVRx HMD[a]—3D applied VR[b] [35]	Chronic lower back pain	3D VR delivery	No	No
Head	Oculus Rift HMD [37]	Chronic neuropathic pain after spinal cord injury	3D VR delivery	No	Yes
Head	Google Glass [57]	Parkinson disease	Augmented reality	No	Yes
Multiple	Oculus Rift HMD and OptiTrack V120—motion-tracking devices [60]	Cerebral palsy and developmental dyspraxia	VR delivery and linear path tracking	No	No
Multiple	Gamepad (wearable 6-inertial sensor system) [52]	Parkinson disease	Visual feedback on movement	No	No
Multiple	Digital medicine offering—ingestible sensor and smart patch [43]	Hypertension and diabetes mellitus	Medication ingestion adherence	Yes	No
Waist	Pedometer (unspecified) [34]	Multiple chronic diseases	Step count	No	Yes
Waist	Yamax DigiWalker CW-701 pedometer [33]	Chronic lower back pain	Step count	No	Yes
Waist	Fitbit Zip [38,40,56]	COPD[c], rheumatoid arthritis, and type 2 diabetes	Step count and calories	No	Yes
Waist	Omron HJ-321 pedometer [48]	COPD	Step count	No	Yes
Waist	Omron HJ-720ITC pedometer [31]	Juvenile idiopathic arthritis	Step count	No	Yes
Waist	Coffee WALKIE+Dv3 pedometer [54]	Metabolic syndrome	Step count	No	Yes
Waist	Omron HJ-112 [45,59]	Obesity with multimorbidity	Step count	No	Yes
Wrist	Fitbit Charge [46]	Type 2 diabetes mellitus	Step count and calories	Yes	Yes
Wrist	Fitbit Charge HR [39]	Advanced liver disease	Step count, heart rate, and calories	Yes	Yes
Wrist	Fitbit Flex [48]	Osteoarthritis (knee)	Step count, distance, calories, and sleep stage estimation	Yes	Yes
Wrist	Heart Rate Smart Wristband GSH405-B6 [61]	Chronic kidney disease	Step count, calories, and sleep stage estimation	Yes	Yes
Wrist	Fitbit Charge 2 [58]	Osteoarthritis	Step count, calories, heart rate, floors climbed	Yes	Yes
Wrist	Fitbit Flex 2 [36]	Rheumatoid arthritis and systemic lupus erythematosus	Step count, calories, and sleep stage estimation	Yes	Yes
Wrist	Nike+ FuelBand [51]	Peripheral vascular disease	Step count and calories	Yes	Yes
Wrist	Samsung Charm [32]	Obstructive sleep apnea	Step count, calories, and distance	Yes	Yes

[a]HMD: head-mounted display.

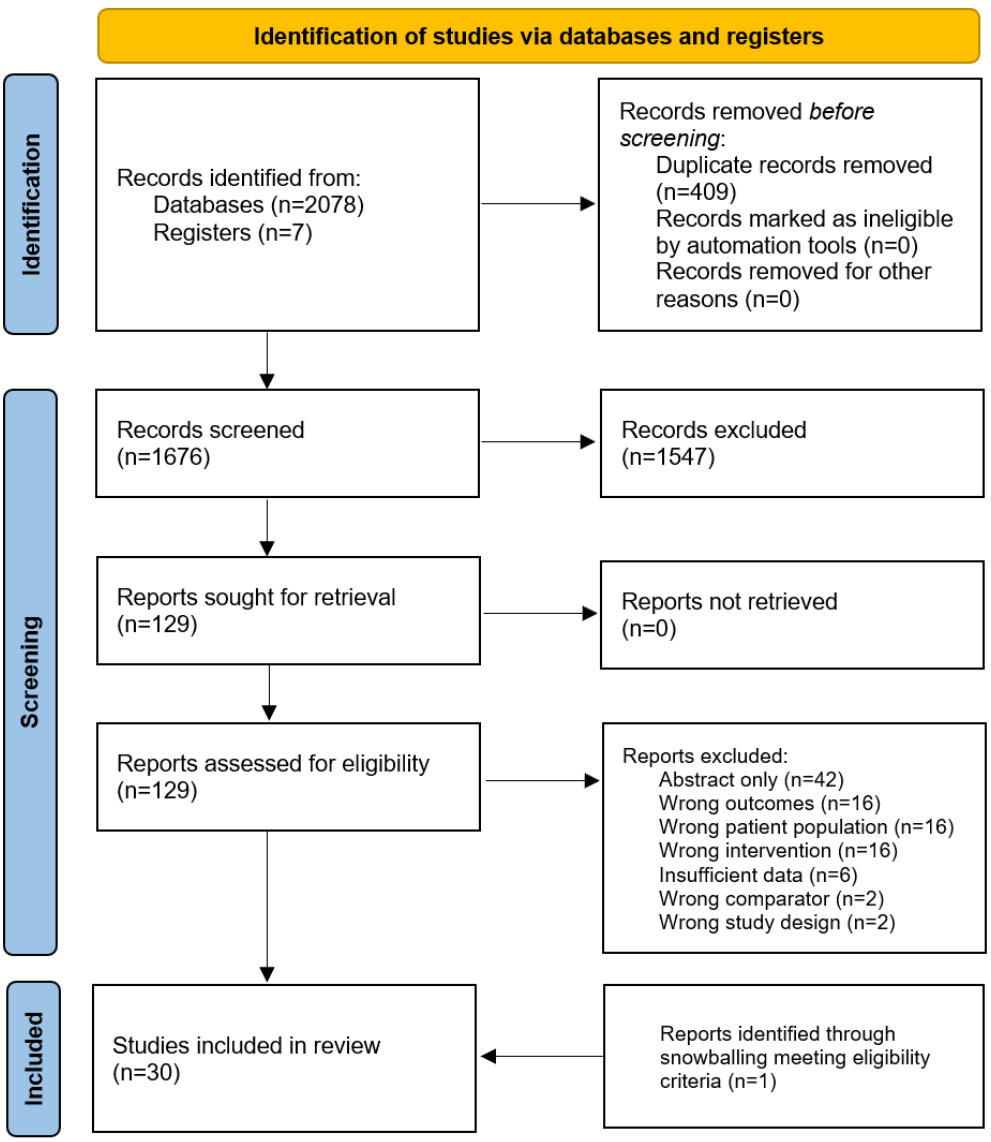
[b]VR: virtual reality.

[c]COPD: chronic obstructive pulmonary disease.

Study Selection

The study selection sequence is outlined in a PRISMA flowchart, which is presented in Figure 1. Our search yielded 2078 articles, with a further 7 articles identified through snowballing. From a total of 2085 articles, 409 (19.62%) were identified as duplicates and were thus removed, and a further 1576 (75.59%) studies were excluded after screening abstracts between July and August 2021, leaving 129 (6.19%) studies assessed for eligibility via full-text review, of which 99 (4.75%) were excluded. One study was identified through snowballing. A total of 31 studies met the inclusion criteria. Following peer review, 1 study was removed, leaving 30 studies included in our qualitative synthesis.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart outlining the study selection sequence.



Study Characteristics

Multimedia Appendix 5 [30-52,54-59] presents the characteristics of all included studies. A total of 2446 participants across 9 countries were included. These participants had a mean age ranging from 10.1 to 74.4 years and 56.42% (1380/2446) were female. Of the 30 studies, 2 (7%) targeted a pediatric population (aged<18 years [30,31], 21 (70%) studied adults with a mean age between 40 and 65 years [32-49,54,55,61], and 6 (20%) evaluated individuals aged >65 years [50-52,56-58]. Approximately 50% (15/30) of studies recruited participants from specialist tertiary clinics, with the remaining 50% (15/30) recruiting participants in the community (eg, primary care settings and rehabilitation centers). Of the 30 studies, 24 (80%) were randomized, with the remaining 6 (20%) using a nonrandomized methodology. All randomized studies were subject to a risk of bias in the blinding of participants because of the nature of the wearable intervention being present or absent, with a notable exception being the use of sham virtual reality (VR) headsets in a study, which used identical hardware with different software with which the user interacted [35].

Chronic Disease Management

Within the literature synthesized, wearables and their influence on health care outcomes in 8 disease systems were studied across 18 chronic diseases. There were predominantly mixed findings within the studies included in this review, which are summarized in the following sections.

Stroke and Neurological Disease

Stroke

Lin et al [41] explored the capability of inertial measurement units (IMUs) in detecting full-body human motion in 20 participants recovering from stroke and assessed the effect of physical activity sessions 3 times per week on upper limb neurological recovery guided by IMUs compared with conventional rehabilitation across 6 weeks. In this study, all 5 Fugl-Meyer assessment (an outcome measure for sensorimotor stroke recovery [62]) subscores improved in both arms, with the deviation angle of shoulder extension rotation during shoulder abduction substantially improving in the IMU group (P=.02) but not in the control group.

Neurological Disease

Approximately 10% (3/30) of studies [50,52,57] explored the effect of a range of wearable systems on motor assessment scoring, balance, self-selected gait speed, and symptom severity scoring in Parkinson disease. These were all perceived as acceptable and enjoyable to use. However, the only positive outcome was the influence of a wearable 6-inertial system on balance and self-selected gait speed when compared with conventional physiotherapy [52]. VR [30,37] and AR [57] systems had mixed results on their desired primary outcomes, although a positive effect was observed in 3D VR interfaces on chronic neuropathic pain after spinal cord injury when compared with sham VR. The reviewed study on the impact of VR on neuropathic pain was limited in size, with only 17 participants being recruited. However, a similar and larger study that incorporated 188 participants demonstrated a positive impact on symptoms in people with chronic lower back pain [35].

Rheumatological and Musculoskeletal Disease

Chronic Lower Back Pain

Chronic lower back pain was studied in 10% (3/30) of publications, with a total of 430 adult participants. Both Amorim et al [42] and Lang et al [33] used commercially available wearables to observe their effects on care-seeking episodes [42] and perceived disability [33], respectively, when compared with usual care. They found no influence on the studied primary outcome. Amorim et al [42] also observed a statistically significant increase of 183.1 min/week in walking time using a Fitbit wearable. However, this did not affect the number of care-seeking episodes in this group. In contrast, Garcia et al [35] studied subjective pain scoring following the implementation of a 3D VR headset (incorporating cognitive behavioral therapy and mindfulness practices) compared with the use of a 2D sham VR system. They found a positive influence of the 3D VR system on subjective pain scoring in addition to secondary outcome measures of pain inference on perceived physical activity, mood, sleep, and stress levels.

Inflammatory Arthritis

Measurement of physical activity was observed in 7% (2/30) of studies that involved people with chronic inflammatory arthritis [31,36], with contrasting results. Blitz et al [31] found a positive influence on the 6-minute walk test (6MWT) in adolescents with juvenile idiopathic arthritis with lower extremity involvement through the use of pedometers and guided education when compared with a pedometer group without education. Li et al [36] found no effect on the moderate to vigorous activity time in people with either rheumatoid arthritis or systemic lupus erythematosus when using a Fitbit Flex 2 compared with people who received usual care. The use of a pedometer coupled with guided education was demonstrated to positively affect subjective fatigue in rheumatoid arthritis in a study by Katz et al [38], who also noted a statistically significant increase in step count in the pedometer group.

Osteoarthritis

Both Zaslavsky et al [58] and Smith et al [47] studied the effect of Fitbit wearables on both sleep quality [58] and exercise capacity (6MWT) [63] in osteoarthritis. Both interventions involved the use of a Fitbit device combined with motivational outputs based on the data provided by the wearable. The results were mixed; Zaslavsky et al [58] found a positive effect of Fitbit-guided exercise on subjective sleep quality, whereas Smith et al [47] did not demonstrate a difference in the 6MWT between the Fitbit and non-Fitbit groups. Sleep quality outcomes were similarly mixed, with improved subjective scoring of sleep quality. However, the objective sleep quality assessed using wrist actigraphy was not affected.

Respiratory Disease

COPD Management

Pulmonary rehabilitation (PR) is a critical component of the nonpharmacological management of COPD [63]. Approximately 7% (2/30) of publications studied the effects of pedometer-guided exercise on adherence to exercise targets [56] and the 6MWT [48] in 70 participants with stable COPD. Ward et al [56] demonstrated 53% adherence to prescribed step count targets with a 20% increase in total step count from week 1 to week 6 of exercise and observed statistically significant improvements in subjective dyspnea, emotional functioning, and disease mastery. Widyastuti et al [48] noted an improvement in the 6MWT using a pedometer from the start of PR to completion; however, this was not greater than the improvements noted in the control group.

Obstructive Sleep Apnea

A randomized, 3-armed study performed by Kim et al [32] recruited 60 individuals with clinician-diagnosed obstructive sleep apnea to assess the effects of a mobile app (MyHealthKeeper) and wearable (Samsung Charm) on weight reduction across 4 weeks compared with an app-only group and an education-only group. They observed significant weight loss in both the app and wearable group (mean weight loss 1.4 kg; P=.02) and app-only group (mean weight loss 2.0 kg; P=.02), which did not translate into secondary outcome measures addressing symptom scoring (snoring frequency, daytime sleepiness, and witnessed apnea).

Cardiovascular Disease: Ischemic Heart Disease

Maddison et al [49] performed a randomized pilot study in 2019 comparing cardiac telerehabilitation with center-based programs for adults with coronary heart disease. Participants were randomized to either a telerehabilitation group—comprising

exercise coaching and monitoring using a chest-worn sensor—or a conventional cardiac rehabilitation group across 12 weeks. The chest-worn sensor allowed real-time monitoring of heart rate, respiratory rate, electrocardiogram, and accelerometry during rehabilitation. The primary outcome measure studied was maximal oxygen consumption (VO_2 max), which refers to the maximum amount of oxygen that an individual can use during maximal exercise and is used as a marker of cardiovascular fitness. VO_2 max was comparable in both groups at 12 weeks, demonstrating noninferiority of telerehabilitation compared with center-based rehabilitation (adjusted mean difference VO_2 max 0.51, 95% CI –0.97 to –1.98 ml/kg/min; *P*=.48).

Endocrine Disease: Type 2 Diabetes Mellitus

Abbott et al [44] studied an intelligent wearable insole (SurroSense Dx) and its ability to prevent diabetic foot ulceration in 58 people with diabetes who recovered from prior foot ulceration. The insole detects high-pressure areas and, in the intervention group, feeds this information back via an integrated app to offload pressure on the affected area. The study did not demonstrate a reduction in the number of diabetic foot ulcer episodes between the groups, despite good adherence to the technology use.

The influence of step count and goal setting on hemoglobin A_{1c} (HbA_{1c}) levels in type 2 diabetes mellitus was studied by both Kooiman et al [40] (using a Fitbit Zip pedometer and web-based self-tracking program) and Lystrup et al [46] (using a Fitbit Charge smartwatch and web-based leaderboards). Both studies reported no significant differences in HbA_{1c} levels after these interventions.

A multifaceted wearable system was studied by Frias et al [43] for people with both diabetes and uncontrolled hypertension comprising an ingestible sensor and an accompanying patch used to detect adherence to prescribed medication in a digital medicine offering. They observed a significant reduction in systolic blood pressure at 4 weeks as the primary outcome of the study, as well as a reduction in systolic blood pressure at 12 weeks, HbA_{1c}, fasting blood glucose, and serum low-density lipoprotein as secondary outcome measures.

Obesity and Metabolic Syndrome

Takahashi et al [45] performed an RCT to determine the effect of pedometer use and behavioral goal setting compared with counseling on exercise and nutrition, step count, gait speed, and grip strength in adults who were overweight and obese with multimorbidities (defined as having>6 comorbid medical conditions). There was no significant improvement in step count between the groups, and although the pedometer and goal-setting groups demonstrated statistically higher grip strength (1 kg; *P*=.01) at 4 months, the clinical importance of this improvement in strength is uncertain. A secondary analysis of this study [59] found no significant differences in within-group weight loss.

Huh et al [54] explored the potential of applying a Coffee WALKIE pedometer with an accompanying mobile app for metabolic syndrome management. They recruited 53 participants with metabolic syndrome and observed the daily step count, calorie expenditure, and proportion of resolved cases of metabolic syndrome following a 12-week study period. Only 20 participants completed the study, and 32 reported communication issues between the wearable and mobile apps, thus leading to withdrawal; however, they observed a mean reduction in systolic blood pressure (mean percentage reduction of 6.71%) and diastolic blood pressure (mean percentage reduction of 7.98%) leading to resolution of metabolic syndrome in 9 participants (*P*=.02).

Chronic Kidney Disease

Li et al [61] studied the ability of a health management platform, incorporating a smart wristband (Heart Rate Smart Wristband GSH405-B6) and accompanying app (WowGoHealth app) to improve participants' self-management abilities and delay the progression of renal decline in chronic kidney disease. All 60 participants were adults, had clinician-diagnosed chronic kidney disease (stages 1 to 4), and were randomized into either the health management platform group or usual care group for a period of 90 days. The intervention group had significantly higher self-efficacy and self-management scores at the end of the study period. The mean kidney disease–related quality of life scores were also significantly higher in the intervention group than in the control group. This translated into reduced rates of renal decline, with a significantly slower decline in estimated glomerular filtration rate observed in the intervention group (–0.56 mL/min/1.73 m^2) than control (–4.58 mL/min/1.73 m^2).

Liver Disease

Chen et al [39] evaluated the impact of a home-based physical activity program on physical fitness using personal activity trackers (Fitbit Charge HR) to remotely monitor and guide exercise efforts in people with advanced liver disease. All 20 participants were provided with a high protein and amino acid diet as a separate exploratory outcome of the study. Physical fitness was assessed using the 6MWT, and computed tomography–based anthropometry was used to assess skeletal muscle volume. The study found a statistically significant improvement in the 6MWT in the home-based physical activity program group versus the control group, with a between-group walking distance difference of 151 m. This did not translate into differences in the computed tomography–based anthropometrics or quality of life outcome measures.

Peripheral Vascular Disease

Normahani et al [51] set out to determine whether the use of a feedback-enabled wearable could improve walking distance and quality of life in people with peripheral vascular disease and intermittent claudication. A total of 37 participants were randomized into an intervention group to use a Nike+ FuelBand with access to data via mobile device pairing and a computer or a control group who received usual care. After 6 months, participants in the Nike+ FuelBand group almost doubled their median maximal walking distance (MWD) from baseline (178 m vs 80 m), and this was sustained at 12 months. Statistically significant improvements were also observed in the distance to the onset of claudication, with this distance improving by 75 m from the baseline at 6 months (112 m vs 40 m). Participants in

the control group did not display improvements in MWD or distance to claudication onset. An improvement in Vascular Quality of Life Questionnaire scores at 12 months was also observed in the Nike+ FuelBand group, which correlated with improvements in the MWD and claudication distance.

Theoretical Aspects of Wearables in Improving Health Care Outcomes

Of the 30 studies, 15 (50%) found a positive impact of wearables on the primary outcome studied. These findings were observed in studies involving multiple chronic disease systems and using multiple wearables. Wearables can be used to deliver nonpharmacological therapy to improve subjective pain scoring, as demonstrated by both Austin et al [37] and Garcia et al [35] in using 3D VR systems to deliver cognitive behavioral therapy and mindfulness practices to users with chronic pain. Tunur et al [59] found a lack of effect in delivering therapy through AR-based dance interventions in Parkinson disease, although the outcome studied was motor assessment, and only 7 participants were enrolled in the study.

Approximately 37% (11/30) of studies observed the capability of wearables to encourage and improve physical activity. There was no uniform measurement system to quantify physical activity across the included studies, with the 6MWT [31,39,47,48], 10-minute walk test [52], minutes of moderate to vigorous physical activity time [34,36], step count [45], MWD [51], VO_2 max [49], and adherence to prescribed walking therapy [56] used to measure exercise capacity. A positive impact was observed in 17% (5/30) of studies assessing the influence of wearables on physical activity. There did not appear to be a strong relationship between wearables and improvement in physical activity in these studies.

Approximately 20% (6/30) of studies [33,35,37,38,58,61] used subjective scoring systems to assess the influence of the studied wearables on primary outcome measures. Of these 6 studies, 5 (83%) [35,37,38,58,61] found a positive effect on the primary outcome across multiple chronic diseases. Given the potential for wearables to empower users with additional health-related data, subjective scoring systems may improve significantly using wearables, which may be reflected in health-related quality of life assessments. However, further research is required in this area.

Associations Between Wearables and Outcome Measures

Several associations were noted between the types of wearables used and the studied outcomes. One such observation was the positive effect of 3D VR systems on pain scoring in 2 distinct chronic pain syndromes [35,37], in which both studies used sham VR as the comparator. Both publications reported minimal side effects of using these immersive systems, with specialized programs used to counter the possible effects of cybersickness (motion sickness specific to the use of VR headsets). The use of either pedometers or smartbands (eg, Fitbit devices) had mixed effects on exercise capacity, with 10% (3/30) of studies [31,39,51] reporting improvements in walking distance using these devices and 7% (2/30) of studies [48,63] finding no influence on this outcome. Mixed results were also observed when studying the influence of either pedometers or smartbands on weight. Kim et al [32] found a significant reduction in BMI through the use of a Samsung Charm fitness tracker in people with obstructive sleep apnea. This is in contrast with the findings of both Lystrup et al [46] (using a Fitbit Charge) and Takahashi et al [63] (using a pedometer).

Mapping Outcomes to the Quadruple Aim of Health Care

A total of 155 outcome measures were studied within the included studies. Of these 155 outcomes, 139 (89.7%) addressed the *health care outcomes* component of the Quadruple Aim, with 12 (7.7%) representing the *patient experience* of using wearables. Approximately 1.3% (2/155) of outcome measures evaluated the *clinician experiences* of wearables, with the remaining 1.3% (2/155) addressing *cost*. Within the health care outcomes, the most frequently studied included pain (11/155, 7.1%), quality of life (7/155, 4.5%), step count (7/155, 4.5%), and physical function (5/155, 3.2%). Approximately 7% (2/30) of studies explored the impact of wearables on patient experience as their primary outcome [51,57]. All outcome measures mapped to the patient experience included acceptance of the technology [57], adherence to wearing the device [36,44], compliance [50], engagement [35], presence (using VR) [37], satisfaction [31,35] and usability [50]. A summary of all outcomes in the included studies is presented in Multimedia Appendix 6 [30-52,54-59].

There were no studies in which the primary outcome could be classified as addressing either the clinician experience or cost of health care provision, with only 4 secondary or exploratory outcomes representing these facets of the Quadruple Aim. In the management of Parkinson disease, it was noted that when using data generated by a finger-worn wearable, the consultation time spent with a neurologist was significantly reduced when compared with an in-person clinical assessment, although this was reported with a statistical statement of inequality (range 29-45 minutes vs 45-60 minutes; $P<.05$) [50]. Widyastuti et al [48] performed a cost-effectiveness analysis of pedometer-guided PR in COPD compared with conventional center-based PR and found that the cost of establishing pedometer-guided care was significantly cheaper than that of conventional PR, with a mean cost-saving of €76.3 (US $80.30) per patient. Maddison et al [49] also performed a cost-effectiveness analysis of telerehabilitation versus center-based rehabilitation, reporting a per capita program delivery difference of £1185 (US $1247.10; $P=.02$); however, hospital service use costs were not significantly different ($P=.20$).

Robust Assessment of the Narrative Qualitative Synthesis

A critical reflection of the synthesis process was performed [64]. The methodology used for this synthesis was aligned with the PRISMA guidelines and registered with PROSPERO, and data extraction was performed based on 6 main themes. It was believed that this enabled researchers to answer the predefined research question. Although a wide range of chronic diseases and wearables were studied and included in our synthesis, certain patient populations were not represented, including those

with underlying mental health conditions, chronic skin conditions, or malignancy. This may reflect a limitation of our search strategy or may equally reflect a lack of understanding of the use of wearables for certain disease groups.

The evidence used was not subject to major bias. Most publications involved a relatively small number of participants, and only 80% (24/30) of the studies were randomized. No assumptions were made regarding adherence, acceptability, intended use, or anticipated outcomes in the included studies. No discrepancies were identified within the included studies, and focus areas of future research are highlighted in the *Discussion* section. The strict search criteria precluded the identification of ongoing studies. However, by reviewing the many trial protocols in our abstract screening and review of the gray literature, it is anticipated that the rate of technological advancement and interest in the field will strengthen the existing evidence base in wearables and health care outcomes in chronic disease management. The aspects that may influence the implementation of wearables in existing health care platforms are highlighted in the *Discussion* section.

Discussion

Principal Findings

A total of 30 studies published between 2016 and 2021 investigated the effectiveness of wearables in improving health care outcomes in individuals with chronic diseases. Our systematic review found both positive and neutral results when studying the influence of wearables on health care outcomes in chronic disease. Encouragingly, none of the identified studies demonstrated a negative impact of wearables on outcomes, such as adverse effects, treatment burden, or the inability of study participants to effectively use the prescribed technology because of poor inadequate digital literacy.

We identified several studies demonstrating the positive influence of the studied wearables on primary health care outcomes. One such example is the positive influence of guided exercise prescription using the Nike+ FuelBand on peripheral vascular disease with intermittent claudication [51]. Approximately 7% (2/30) of RCTs found a significant improvement in quality of life indices [51,61], which correlated with either greater self-efficacy in chronic disease management [61] or functional capacity [51].

A key finding was the lack of a clear association between the use of a particular wearable and its acceptability within a chronic disease population. Most health apps synchronizing with wearables, including those studied in this review, are disease focused and provide information on the studied conditions, such as pressure points in diabetic foot ulceration. Although this does provide clinically useful information, it is unclear whether patients use this information effectively to guide their self-management or even find it useful. We also have limited knowledge of the acceptability of wearables to patients with chronic disease or whether there are any barriers to implementation within specific chronic disease groups. Supporting mobile apps should be designed using a patient-centered approach, incorporating personalized advice and recommendations from the health data provided by the accompanying wearable. This may unlock the potential of wearables in chronic diseases.

To rationalize the incorporation of wearables into existing health care models, it is important to demonstrate cost-effectiveness; however, economic outcomes were not incorporated into the publications identified in this review. Wearables enable remote monitoring, which has its own economic advantages. One of the studies involving patients with implantable cardiac defibrillators found that remote monitoring reduced health care costs by 25% [65]. Transferring even a small proportion of *center-based* monitoring services established for people with chronic disease remotely through the use of wearables has the potential to substantially reduce the cost of care delivery in many chronic disease settings. Robust economic analyses incorporating the number needed to treat analyses should be inherent to future studies on wearable technologies.

Implications for Clinical Practice

There are several challenges in the implementation of wearables in health care. Wearables were initially designed for the health and fitness industry and are not subject to the regulatory standards required for medical equipment. Only a handful of wearable technologies have been approved by the US Food and Drug Administration, such as the Apple Watch Series 6 electrocardiogram app for detecting atrial fibrillation [66]. Most commercially available wearables are classified as nonmedical devices, and although some software within these are accurate, they cannot support clinical decision-making without legal regulation [67]. We noted that 73% (22/30) of studies included in this systematic review used commercially available wearable technology that is not currently legislated under either the US Food and Drug Administration or the Therapeutic Goods Administration (Australia); therefore, this may represent a limitation in applicability to health care settings. Wearables can also collect many different aspects of user information using sensor technology, which requires stringent data security and privacy processes. The accuracy of recently implemented software in wearables is variable and must be recognized to prevent avoidable misdiagnosis and unnecessary patient-clinician interfaces and minimize patient anxiety [68]. Until these challenges are clearly addressed, it is difficult to envision the safe implementation of wearables in existing health care models.

There is a high degree of variance in the chronic diseases studied and in the wearables used, which limits the ability to provide a strong evidence base to support the use of a specific wearable to treat a specific cause. Further research needs to focus on the impact of a specific wearable for a specific chronic disease to generate evidence for its use, especially given the multiple capabilities of most wearables. In addition, there may be several applications within the studied wearables for which only one wearable is particularly effective for a chronic disease group. Understanding these capabilities through targeted research is essential for the implementation of wearables in chronic disease management.

A finding from this systematic review was the lack of studies focusing on the clinician experience of wearables in health care.

Without a deep understanding of the perceived benefits and risks of wearables in chronic disease management from the clinician's perspective, it is difficult to envision their integration into clinical practice. Challenges to more established digital health transformative initiatives, such as introducing electronic health records, likely exist for the use of wearables in health care. These include changes to clinical workflows, patient safety concerns, learning curves of practitioners in understanding new technology, and challenges in integrating systems [69]. Interdisciplinary teams, including clinicians, data scientists, and experts in artificial intelligence, will be required to shape the future of wearables in health care because of the limited familiarity of clinicians with big data and artificial intelligence [70]. Alongside observational and randomized controlled quantitative analyses, qualitative research into the clinician's perspective will prove invaluable in optimizing the use of wearables in health care.

Strengths and Limitations of This Review

There were several strengths of our review. Given the rapid shift to care decentralization amidst a global pandemic [71], this review provides a foundational evidence base for the effect of wearables in improving health care outcomes for individuals with chronic diseases. A large number of participants was obtained, encompassing a range of chronic diseases, with a wide mean age range and reasonably equal gender distribution. Approximately 80% (24/30) of studies implemented a randomized controlled design, generating a comparator to assess the effect of the wearable intervention studied, which was predominantly the gold standard of care for the studied chronic disease. Although pedometers have been outdated commercially by newer technology, their inclusion as wearables in this study remains a strength because of the long-term evidence of their ability to encourage physical activity [72].

Our review has several limitations, such as the use of now outdated wearables. All smartwatch brands used in these studies have since been superseded by more advanced devices with greater technical capabilities. Although this may simply reflect the rate of technological advancement, it may also indicate that the studies published in this review may not reflect the capabilities of wearable technology from 2021 onward. It is unknown whether advanced information, including heart rate variability, Firstbeat algorithms such as stress levels and Garmin's Body Battery, and advanced sleep stage estimation may infer real-time information on disease status. The rate of increasing technical capability of wearables justifies the regular systematic review of the literature, given the increasing publication outputs and commercial availability of more advanced wearables.

The interventions used as comparators differed greatly between studies, which made the comparison of the results challenging. The heterogeneity of the participants in each study precluded the ability to perform a meta-analysis, leading to an inability to assess the strength of evidence for a particular wearable on a predefined health care outcome. Although there were no apparent major risks of bias in our risk of bias assessments, this process was conducted by only 1 researcher. In addition, within the included studies, there was a lack of multicenter trials, which we propose should be conducted to increase the statistical robustness. All outcome measures reported were quantitative in nature, which represents a limitation, as qualitative research strengthens the existing evidence base. Given the increasing number of publications in this field, it is highly likely that a meta-analysis will be feasible for future systematic reviews of this nature. In future studies, all 4 end points of the Quadruple Aim should be represented to aid implementation in health care systems. We believe that further research is worth conducting to strengthen the evidence regarding wearables in chronic disease management.

Conclusions

Our systematic review did not find a clear role of wearables in improving health care outcomes in chronic disease. However, wearables are becoming increasingly popular within the community, and as research and development in wearable technology progress, it is anticipated that these devices will play an increasing role in supporting healthy lifestyle modifications for their users. More research is required to ascertain a clear causality between wearables and health care outcomes, as defined by the Quadruple Aim, for people with chronic disease. As the evidence base for the use of wearables in chronic disease management is strengthened, further challenges will need to be overcome to allow widespread adoption in the health care setting, including stringent regulatory approval, data privacy and confidentiality, and software accuracy.

Acknowledgments

The lead author (GM) was supported by The University of Queensland and the Digital Health Cooperative Research Centre, Sydney Startup Hub, Sydney, in completing his PhD and would like to acknowledge this support. GM would like to acknowledge Christine Dalais, Liaison Librarian, University of Queensland, for her contribution to the search strategy implemented in this review.

Authors' Contributions

All listed authors have reviewed and contributed to this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 290 KB - jmir_v24i7e36690_app1.docx]

Multimedia Appendix 2
Search strings for systematic review.
[DOCX File , 37 KB - jmir_v24i7e36690_app2.docx]

Multimedia Appendix 3
Risk of bias assessment for randomized studies.
[DOCX File , 17 KB - jmir_v24i7e36690_app3.docx]

Multimedia Appendix 4
Risk of bias assessment for nonrandomized studies.
[DOCX File , 29 KB - jmir_v24i7e36690_app4.docx]

Multimedia Appendix 5
Summary of articles included in systematic review classified by disease group.
[DOCX File , 70 KB - jmir_v24i7e36690_app5.docx]

Multimedia Appendix 6
Outcome measures within included studies.
[DOCX File , 50 KB - jmir_v24i7e36690_app6.docx]

References

1. GBD 2017 Causes of Death Collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1736-1788 [FREE Full text] [doi: 10.1016/S0140-6736(18)32203-7] [Medline: 30496103]
2. Noncommunicable diseases. World Health Organisation. 2016. URL: http://www.who.int/topics/noncommunicable_diseases/en/ [accessed 2021-11-03]
3. Fried TR, Bradley EH, Williams CS, Tinetti ME. Functional disability and health care expenditures for older persons. Arch Intern Med 2001 Nov 26;161(21):2602-2607. [doi: 10.1001/archinte.161.21.2602] [Medline: 11718592]
4. Kovacs FM, Abraira V, Zamora J, Teresa Gil del Real M, Llobera J, Fernández C, Kovacs-Atención Primaria Group. Correlation between pain, disability, and quality of life in patients with common low back pain. Spine (Phila Pa 1976) 2004 Jan 15;29(2):206-210. [doi: 10.1097/01.BRS.0000107235.47465.08] [Medline: 14722416]
5. Australia's health 2014 - Cat. no: AUS 178. Australian Institute of Health and Welfare. 2014 Jun 25. URL: https://www.aihw.gov.au/reports/australias-health/australias-health-2014/overview [accessed 2021-03-26]
6. Grady PA, Gough LL. Self-management: a comprehensive approach to management of chronic conditions. Am J Public Health 2014 Aug;104(8):e25-e31. [doi: 10.2105/AJPH.2014.302041] [Medline: 24922170]
7. 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]
8. Chodosh J, Morton SC, Mojica W, Maglione M, Suttorp MJ, Hilton L, et al. Meta-analysis: chronic disease self-management programs for older adults. Ann Intern Med 2005 Sep 20;143(6):427-438. [doi: 10.7326/0003-4819-143-6-200509200-00007] [Medline: 16172441]
9. Jordan JE, Briggs AM, Brand CA, Osborne RH. Enhancing patient engagement in chronic disease self-management support initiatives in Australia: the need for an integrated approach. Med J Aust 2008 Nov 17;189(S10):S9-13. [doi: 10.5694/j.1326-5377.2008.tb02202.x] [Medline: 19143585]
10. National Strategic Framework for Chronic Conditions. Australian Health Ministers' Advisory Council. Canberra, Australia: Australian Government; 2017. URL: https://www.health.gov.au/resources/publications/national-strategic-framework-for-chronic-conditions [accessed 2021-03-26]
11. Digital Health Network Research Capability. Global Change Institute - The University of Queensland. 2021. URL: https://gci.uq.edu.au/DHRN-researchers [accessed 2021-04-01]
12. Wu RC, Ginsburg S, Son T, Gershon AS. Using wearables and self-management apps in patients with COPD: a qualitative study. ERJ Open Res 2019 Jul;5(3):00036-2019 [FREE Full text] [doi: 10.1183/23120541.00036-2019] [Medline: 31528634]
13. Kamei T, Kanamori T, Yamamoto Y, Edirippulige S. The use of wearable devices in chronic disease management to enhance adherence and improve telehealth outcomes: a systematic review and meta-analysis. J Telemed Telecare 2022 Jun;28(5):342-359. [doi: 10.1177/1357633X20937573] [Medline: 32819184]

14. 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]
15. Bonato P. Advances in wearable technology and its medical applications. In: Proceedings of the 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology. 2010 Presented at: IEMBS '10; August 31-September 4, 2010; Buenos Aires, Argentina p. 2021-2024. [doi: 10.1109/iembs.2010.5628037]
16. Dobkin BH, Dorsch A. The promise of mHealth: daily activity monitoring and outcome assessments by wearable sensors. Neurorehabil Neural Repair 2011;25(9):788-798 [FREE Full text] [doi: 10.1177/1545968311425908] [Medline: 21989632]
17. Sharma A, Badea M, Tiwari S, Marty JL. Wearable biosensors: an alternative and practical approach in healthcare and disease monitoring. Molecules 2021 Feb 01;26(3):748 [FREE Full text] [doi: 10.3390/molecules26030748] [Medline: 33535493]
18. Brickwood KJ, 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]
19. Vijayan V, Connolly JP, Condell J, McKelvey N, Gardiner P. Review of wearable devices and data collection considerations for connected health. Sensors (Basel) 2021 Aug 19;21(16):5589 [FREE Full text] [doi: 10.3390/s21165589] [Medline: 34451032]
20. Hickey AM, Freedson PS. Utility of consumer physical activity trackers as an intervention tool in cardiovascular disease prevention and treatment. Prog Cardiovasc Dis 2016;58(6):613-619. [doi: 10.1016/j.pcad.2016.02.006] [Medline: 26943981]
21. Cancela J, Pastorino M, Tzallas AT, Tsipouras MG, Rigas G, Arredondo MT, et al. Wearability assessment of a wearable system for Parkinson's disease remote monitoring based on a body area network of sensors. Sensors (Basel) 2014 Sep 16;14(9):17235-17255 [FREE Full text] [doi: 10.3390/s140917235] [Medline: 25230307]
22. Ummels D, Beekman E, Theunissen K, Braun S, Beurskens AJ. Counting steps in activities of daily living in people with a chronic disease using nine commercially available fitness trackers: cross-sectional validity study. JMIR Mhealth Uhealth 2018 Apr 02;6(4):e70 [FREE Full text] [doi: 10.2196/mhealth.8524] [Medline: 29610110]
23. Howse E, Rychetnik L, Marks L, Wilson A. What does the future hold for chronic disease prevention research? Aust N Z J Public Health 2020 Oct;44(5):336-340. [doi: 10.1111/1753-6405.13028] [Medline: 32865859]
24. Kristoffersson A, Lindén M. A systematic review on the use of wearable body sensors for health monitoring: a qualitative synthesis. Sensors (Basel) 2020 Mar 09;20(5):1502 [FREE Full text] [doi: 10.3390/s20051502] [Medline: 32182907]
25. 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]
26. Perez AJ, Zeadally S. Recent advances in wearable sensing technologies. Sensors (Basel) 2021 Oct 14;21(20):6828 [FREE Full text] [doi: 10.3390/s21206828] [Medline: 34696040]
27. Chronic diseases. Australian Institute of Health and Welfare. 2016. URL: http://www.aihw.gov.au/chronic-diseases/ [accessed 2021-03-26]
28. Covidence. URL: https://www.covidence.org/home [accessed 2021-07-01]
29. Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions Version 6.2. London, UK: Cochrane Collaboration; 2021.
30. Bortone I, Barsotti M, Leonardis D, Crecchi A, Tozzini A, Bonfiglio L, et al. Immersive Virtual Environments and Wearable Haptic Devices in rehabilitation of children with neuromotor impairments: a single-blind randomized controlled crossover pilot study. J Neuroeng Rehabil 2020 Oct 28;17(1):144 [FREE Full text] [doi: 10.1186/s12984-020-00771-6] [Medline: 33115487]
31. Blitz J, Stern SM, Marzan K. Do pedometers with or without education on exercise increase functional walking capacity and physical activity level in adolescents with juvenile idiopathic arthritis? Physiother Theory Pract 2018 May;34(5):359-366. [doi: 10.1080/09593985.2017.1401189] [Medline: 29131689]
32. Kim JW, Ryu B, Cho S, Heo E, Kim Y, Lee J, et al. Impact of personal health records and wearables on health outcomes and patient response: three-arm randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 04;7(1):e12070 [FREE Full text] [doi: 10.2196/12070] [Medline: 30609978]
33. Lang AE, Hendrick PA, Clay L, Mondal P, Trask CM, Bath B, et al. A randomized controlled trial investigating effects of an individualized pedometer driven walking program on chronic low back pain. BMC Musculoskelet Disord 2021 Feb 19;22(1):206 [FREE Full text] [doi: 10.1186/s12891-021-04060-8] [Medline: 33607979]
34. Taylor A, Taylor RS, Ingram W, Dean SG, Jolly K, Mutrie N, et al. Randomised controlled trial of an augmented exercise referral scheme using web-based behavioural support for inactive adults with chronic health conditions: the e-coachER trial. Br J Sports Med 2021 Apr;55(8):444-450 [FREE Full text] [doi: 10.1136/bjsports-2020-103121] [Medline: 33247001]
35. Garcia LM, Birckhead BJ, Krishnamurthy P, Sackman J, Mackey IG, Louis RG, et al. An 8-week self-administered at-home behavioral skills-based virtual reality program for chronic low back pain: double-blind, randomized, placebo-controlled trial conducted during COVID-19. J Med Internet Res 2021 Feb 22;23(2):e26292 [FREE Full text] [doi: 10.2196/26292] [Medline: 33484240]

36. Li LC, Feehan LM, Xie H, Lu N, Shaw C, Gromala D, et al. Efficacy of a physical activity counseling program with use of a wearable tracker in people with inflammatory arthritis: a randomized controlled trial. Arthritis Care Res (Hoboken) 2020 Dec;72(12):1755-1765. [doi: 10.1002/acr.24199] [Medline: 32248626]
37. Austin PD, Craig A, Middleton JW, Tran Y, Costa DS, Wrigley PJ, et al. The short-term effects of head-mounted virtual-reality on neuropathic pain intensity in people with spinal cord injury pain: a randomised cross-over pilot study. Spinal Cord 2021 Jul;59(7):738-746. [doi: 10.1038/s41393-020-00569-2] [Medline: 33077900]
38. Katz P, Margaretten M, Gregorich S, Trupin L. Physical activity to reduce fatigue in rheumatoid arthritis: a randomized controlled trial. Arthritis Care Res (Hoboken) 2018 Jan;70(1):1-10. [doi: 10.1002/acr.23230] [Medline: 28378441]
39. Chen HW, Ferrando A, White MG, Dennis RA, Xie J, Pauly M, et al. Home-based physical activity and diet intervention to improve physical function in advanced liver disease: a randomized pilot trial. Dig Dis Sci 2020 Nov;65(11):3350-3359. [doi: 10.1007/s10620-019-06034-2] [Medline: 31907774]
40. Kooiman TJ, de Groot M, Hoogenberg K, Krijnen WP, van der Schans CP, Kooy A. Self-tracking of physical activity in people with type 2 diabetes: a randomized controlled trial. Comput Inform Nurs 2018 Jul;36(7):340-349. [doi: 10.1097/CIN.0000000000000443] [Medline: 29742550]
41. Lin LF, Lin YJ, Lin ZH, Chuang LY, Hsu WC, Lin YH. Feasibility and efficacy of wearable devices for upper limb rehabilitation in patients with chronic stroke: a randomized controlled pilot study. Eur J Phys Rehabil Med 2018 Jun;54(3):388-396 [FREE Full text] [doi: 10.23736/S1973-9087.17.04691-3] [Medline: 28627862]
42. Amorim AB, Pappas E, Simic M, Ferreira ML, Jennings M, Tiedemann A, et al. Integrating mobile-health, health coaching, and physical activity to reduce the burden of chronic low back pain trial (IMPACT): a pilot randomised controlled trial. BMC Musculoskelet Disord 2019 Feb 11;20(1):71 [FREE Full text] [doi: 10.1186/s12891-019-2454-y] [Medline: 30744606]
43. Frias J, Virdi N, Raja P, Kim Y, Savage G, Osterberg L. Effectiveness of digital medicines to improve clinical outcomes in patients with uncontrolled hypertension and type 2 diabetes: prospective, open-label, cluster-randomized pilot clinical trial. J Med Internet Res 2017 Jul 11;19(7):e246 [FREE Full text] [doi: 10.2196/jmir.7833] [Medline: 28698169]
44. Abbott CA, Chatwin KE, Foden P, Hasan AN, Sange C, Rajbhandari SM, et al. Innovative intelligent insole system reduces diabetic foot ulcer recurrence at plantar sites: a prospective, randomised, proof-of-concept study. Lancet Digit Health 2019 Oct;1(6):e308-e318 [FREE Full text] [doi: 10.1016/S2589-7500(19)30128-1] [Medline: 33323253]
45. Takahashi PY, Quigg SM, Croghan IT, Schroeder DR, Ebbert JO. Effect of pedometer use and goal setting on walking and functional status in overweight adults with multimorbidity: a crossover clinical trial. Clin Interv Aging 2016 Sep 1;11:1099-1106 [FREE Full text] [doi: 10.2147/CIA.S107626] [Medline: 27621602]
46. Lystrup R, Carlsen D, Sharon DJ, Crawford P. Wearable and interactive technology to share fitness goals results in weight loss but not improved diabetes outcomes. Obes Res Clin Pract 2020;14(5):443-448. [doi: 10.1016/j.orcp.2020.08.006] [Medline: 32891527]
47. Smith WA, Zucker-Levin A, Mihalko WM, Williams M, Loftin M, Gurney JG. A randomized study of exercise and fitness trackers in obese patients after total knee arthroplasty. Orthop Clin North Am 2019 Jan;50(1):35-45. [doi: 10.1016/j.ocl.2018.08.002] [Medline: 30477705]
48. Widyastuti K, Makhabah DN, Setijadi AR, Sutanto YS, Suradi, Ambrosino N. Benefits and costs of home pedometer assisted physical activity in patients with COPD. A preliminary randomized controlled trial. Pulmonology 2018;24(4):211-218 [FREE Full text] [doi: 10.1016/j.pulmoe.2018.01.006] [Medline: 30008335]
49. Maddison R, Rawstorn JC, Stewart RA, Benatar J, Whittaker R, Rolleston A, et al. Effects and costs of real-time cardiac telerehabilitation: randomised controlled non-inferiority trial. Heart 2019 Jan;105(2):122-129 [FREE Full text] [doi: 10.1136/heartjnl-2018-313189] [Medline: 30150328]
50. Heldman DA, Harris DA, Felong T, Andrzejewski KL, Dorsey ER, Giuffrida JP, et al. Telehealth management of Parkinson's disease using wearable sensors: an exploratory study. Digit Biomark 2017 Sep;1(1):43-51 [FREE Full text] [doi: 10.1159/000475801] [Medline: 29725667]
51. Normahani P, Kwasnicki R, Bicknell C, Allen L, Jenkins MP, Gibbs R, et al. Wearable sensor technology efficacy in peripheral vascular disease (wSTEP): a randomized controlled trial. Ann Surg 2018 Dec;268(6):1113-1118. [doi: 10.1097/SLA.0000000000002300] [Medline: 28498233]
52. Carpinella I, Cattaneo D, Bonora G, Bowman T, Martina L, Montesano A, et al. Wearable sensor-based biofeedback training for balance and gait in Parkinson disease: a pilot randomized controlled trial. Arch Phys Med Rehabil 2017 Apr;98(4):622-30.e3. [doi: 10.1016/j.apmr.2016.11.003] [Medline: 27965005]
53. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016 Oct 12;355:i4919 [FREE Full text] [doi: 10.1136/bmj.i4919] [Medline: 27733354]
54. Huh U, Tak YJ, Song S, Chung SW, Sung SM, Lee CW, et al. Feedback on physical activity through a wearable device connected to a mobile phone app in patients with metabolic syndrome: pilot study. JMIR Mhealth Uhealth 2019 Jun 18;7(6):e13381 [FREE Full text] [doi: 10.2196/13381] [Medline: 31215513]
55. Stuart CM, Varatharaj A, Domjan J, Philip S, Galea I, SIMS study group. Physical activity monitoring to assess disability progression in multiple sclerosis. Mult Scler J Exp Transl Clin 2020 Dec 7;6(4):2055217320975185 [FREE Full text] [doi: 10.1177/2055217320975185] [Medline: 33343919]

56. Ward S, Orme M, Zatloukal J, Singh S. Adherence to walking exercise prescription during pulmonary rehabilitation in COPD with a commercial activity monitor: a feasibility trial. BMC Pulm Med 2021 Jan 18;21(1):30 [FREE Full text] [doi: 10.1186/s12890-021-01406-9] [Medline: 33461515]
57. Tunur T, DeBlois A, Yates-Horton E, Rickford K, Columna LA. Augmented reality-based dance intervention for individuals with Parkinson's disease: a pilot study. Disabil Health J 2020 Apr;13(2):100848. [doi: 10.1016/j.dhjo.2019.100848] [Medline: 31679951]
58. Zaslavsky O, Thompson HJ, McCurry SM, Landis CA, Kitsiou S, Ward TM, et al. Use of a wearable technology and motivational interviews to improve sleep in older adults with osteoarthritis and sleep disturbance: a pilot study. Res Gerontol Nurs 2019 Jul 01;12(4):167-173 [FREE Full text] [doi: 10.3928/19404921-20190319-02] [Medline: 30901479]
59. Takahashi PY, Quigg SM, Croghan IT, Schroeder DR, Ebbert JO. SMART goals setting and biometric changes in obese adults with multimorbidity: secondary analysis of a randomized controlled trial. SAGE Open Med 2019 Jun 24;7:2050312119858042 [FREE Full text] [doi: 10.1177/2050312119858042] [Medline: 31258905]
60. Institute of Medicine (US) Committee on Standards for Systematic Reviews of Comparative Effectiveness Research, Eden J, Levit L, Berg A, Morton S. Finding What Works in Health Care: Standards for Systematic Reviews. Washington, DC, USA: National Academies Press (US); 2011.
61. Li WY, Chiu FC, Zeng JK, Li YW, Huang SH, Yeh HC, et al. Mobile health app with social media to support self-management for patients with chronic kidney disease: prospective randomized controlled study. J Med Internet Res 2020 Dec 15;22(12):e19452 [FREE Full text] [doi: 10.2196/19452] [Medline: 33320101]
62. Gladstone DJ, Danells CJ, Black SE. The fugl-meyer assessment of motor recovery after stroke: a critical review of its measurement properties. Neurorehabil Neural Repair 2002 Sep;16(3):232-240. [doi: 10.1177/154596802401105171] [Medline: 12234086]
63. 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;8(2):89-99 [FREE Full text] [doi: 10.1177/1479972310393756] [Medline: 21596892]
64. 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. Version1.1. Lancaster University. 2006. URL: https://tinyurl.com/3ex2n5fx [accessed 2022-04-15]
65. Costa PD, Reis AH, Rodrigues PP. Clinical and economic impact of remote monitoring on the follow-up of patients with implantable electronic cardiovascular devices: an observational study. Telemed J E Health 2013 Feb;19(2):71-80. [doi: 10.1089/tmj.2012.0064] [Medline: 23336734]
66. Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T, Apple Heart Study Investigators. Large-scale assessment of a smartwatch to identify atrial fibrillation. N Engl J Med 2019 Nov 14;381(20):1909-1917 [FREE Full text] [doi: 10.1056/NEJMoa1901183] [Medline: 31722151]
67. Lu L, Zhang J, Xie Y, Gao F, Xu S, Wu X, et al. Wearable health devices in health care: narrative systematic review. JMIR Mhealth Uhealth 2020 Nov 09;8(11):e18907 [FREE Full text] [doi: 10.2196/18907] [Medline: 33164904]
68. Segura Anaya LH, Alsadoon A, Costadopoulos N, Prasad PW. Ethical implications of user perceptions of wearable devices. Sci Eng Ethics 2018 Feb;24(1):1-28. [doi: 10.1007/s11948-017-9872-8] [Medline: 28155094]
69. Sullivan C, Staib A, Ayre S, Daly M, Collins R, Draheim M, et al. Pioneering digital disruption: Australia's first integrated digital tertiary hospital. Med J Aust 2016 Nov 07;205(9):386-389. [doi: 10.5694/mja16.00476] [Medline: 27809727]
70. Majnarić LT, Babič F, O'Sullivan S, Holzinger A. AI and big data in healthcare: towards a more comprehensive research framework for multimorbidity. J Clin Med 2021 Feb 14;10(4):766 [FREE Full text] [doi: 10.3390/jcm10040766] [Medline: 33672914]
71. Annis T, Pleasants S, Hultman G, Lindemann E, Thompson JA, Billecke S, et al. Rapid implementation of a COVID-19 remote patient monitoring program. J Am Med Inform Assoc 2020 Aug 01;27(8):1326-1330 [FREE Full text] [doi: 10.1093/jamia/ocaa097] [Medline: 32392280]
72. Harris T, Kerry SM, Limb ES, Furness C, Wahlich C, Victor CR, et al. Physical activity levels in adults and older adults 3-4 years after pedometer-based walking interventions: long-term follow-up of participants from two randomised controlled trials in UK primary care. PLoS Med 2018 Mar 9;15(3):e1002526 [FREE Full text] [doi: 10.1371/journal.pmed.1002526] [Medline: 29522529]

Abbreviations

6MWT: 6-minute walk test
AR: augmented reality
COPD: chronic obstructive pulmonary disease
HbA_{1c}: hemoglobin A_{1c}
IMU: inertial measurement unit
MeSH: Medical Subject Headings
MWD: maximal walking distance

PR: pulmonary rehabilitation
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register for Systematic Reviews
RCT: randomized controlled trial
VO_2 max: maximal oxygen consumption
VR: virtual reality

Edited by T Leung; submitted 27.01.22; peer-reviewed by J Koetsenruijter, M Bonacini, J Campbell, D Joyner; comments to author 30.03.22; revised version received 19.04.22; accepted 16.05.22; published 01.07.22.

Please cite as:
Mattison G, Canfell O, Forrester D, Dobbins C, Smith D, Töyräs J, Sullivan C
The Influence of Wearables on Health Care Outcomes in Chronic Disease: Systematic Review
J Med Internet Res 2022;24(7):e36690
URL: https://www.jmir.org/2022/7/e36690
doi:10.2196/36690
PMID:35776492

Review

Behavioral Change Factors and Retention in Web-Based Interventions for Informal Caregivers of People Living With Dementia: Scoping Review

Kuan-Ching Wu[1], BSN; Yan Su[1], BSN, MS, PhD; Frances Chu[1], MLIS; Annie T Chen[2], MSIS, PhD; Oleg Zaslavsky[3], MHA, PhD

[1]School of Nursing, University of Washington, Seattle, WA, United States

[2]Department of Biomedical Informatics and Medical Education, University of Washington School of Medicine, Seattle, WA, United States

[3]Biobehavioral Nursing and Health Informatics, University of Washington School of Nursing, Seattle, WA, United States

Corresponding Author:
Kuan-Ching Wu, BSN
School of Nursing
University of Washington
1959 NE Pacific St,
Seattle, WA, 98195
United States
Phone: 1 2065438736
Email: kuancw@uw.edu

Abstract

Background: Web-based interventions aimed at supporting informal caregivers of people living with dementia have the potential to improve caregivers' well-being and psychological health. However, few interventions are widely implemented for this population, and none of the prior reviews have systematically examined the use of behavior change techniques (BCTs), theories, and agents in web-based interventions for informal caregivers of people living with dementia. To better understand this implementation gap, we reviewed the literature to map behavioral factors (BCTs, theories, and agents) deployed in the studies. Furthermore, because there is an emerging consensus that retention could be shaped by participant characteristics and behavioral factors, we explored relationships between these features and retention rates across studies.

Objective: We pursued 3 objectives: to map behavioral factors involved in the web-based interventions for informal caregivers of people living with dementia; to examine the relationship between behavioral change elements and retention in the studies; and to examine the relationship between participant characteristics (gender, age, and spouse or adult children caregiver proportion) and study retention.

Methods: We conducted a literature review using the following keywords and their corresponding Medical Subject Headings terms: dementia, caregivers, and web-based intervention. The time limits were January 1998 to March 2022. Using the BCTv1 taxonomy, which specifies active behavioral components in interventions, 2 coders collected, summarized, and analyzed the frequency distributions of BCTs. Similarly, they abstracted and analyzed participant characteristics, behavior change theories, behavior change agents, and retention rates in the studies.

Results: The average age was 61.5 (SD 7.4) years, and the average proportion of spousal informal caregivers, adult children informal caregivers, and retention rates were 51.2% (SD 24.8%), 44.8% (SD 22%), and 70.4% (SD 17%), respectively. Only 53% (17/32) of the studies used behavior change theories, but 81% (26/32) included behavior change agents. The most common BCTv1 clusters were *shaping knowledge* and *social support*. The median number of *BCTv1 clusters* was 5 (IQR 3). We observed a negative correlation between the proportion of spousal informal caregivers and the retention rate (r=−0.45; P=.02) and between the number of BCTv1 clusters and retention rates (r=−0.47; P=.01). We also found that the proportion of adult children informal caregivers in the study was significantly and positively correlated with the retention rate (r=0.5; P=.03). No other participant characteristics or behavioral factors were associated with retention rates.

Conclusions: We found that almost half of the studies were not informed by behavior change theories. In addition, spousal involvement and a higher number of BCTs were each associated with lower retention rates, while the involvement of adult children caregivers in the study was associated with higher retention. In planning future studies, researchers should consider matching participant characteristics with their intended intervention as the alignment might improve their retention rates.

(J Med Internet Res 2022;24(7):e38595) doi:10.2196/38595

KEYWORDS

dementia; informal caregivers; informal care; caregiving; retention; internet; web-based; behavior; intervention; review; scoping; health intervention; digital health; caregiver; psychological health; cognition; peer support; web-based intervention; taxonomy; aging; gerontology; older adult population; neurological disorder; behavior change technique; BCT; change technique

Introduction

Behavioral Change Interventions for Dementia Caregivers

People living with dementia may have difficulty independently managing their care and typically rely on family and friend caregivers. In fact, 83% of the help provided to older adults with dementia in the United States comes from informal caregivers including family members, friends, or other unpaid caregivers [1]. Given the demand of care involved in dementia, informal caregivers often experience a variety of adverse health complications [1-4]. Compared with caregivers of people living without dementia, informal caregivers of people living with dementia experience 1.5 times higher chances of stroke and a 10% higher occurrence of coronary heart disease, cardiovascular disease, diabetes, and cancer [1,4]. In addition, caregivers of people living with dementia face increased risks of stress, burden, depression, anxiety, and poor quality of life [1-3].

Several behavioral interventions were developed to enhance caregiving knowledge, competencies, and mental health in this population [5-7]. Specifically, cognitive behavioral therapy (CBT) significantly improved depressive symptoms and reduced emotional burden experienced by informal caregivers of people living with dementia [5]. Psychoeducational approaches were also effective in improving caregiving knowledge, well-being, and satisfaction, as well as in reducing caregiver burden, anxiety, and depressive symptoms in caregivers of people living with dementia [6,7]. However, informal caregivers of people living with dementia who are heavily engaged in caregiving tasks or who are at work might not be able to fully participate in the interventions because some behavior programs require face-to-face delivery [8]. Moreover, some specific behavioral interventions are not practical in pandemic settings when in-person contact is discouraged.

The Use of Web-Based Interventions for Dementia Caregivers

Technology is one method to improve access to care by making psychosocial interventions readily accessible. Web-based interventions, which have been used interchangeably with internet- or web-based interventions, are self-guided or therapist-assisted programs that aim to improve knowledge, provide support, care, or treatment to diverse populations with a range of health problems [9]. Web-based interventions that integrate behavioral change interventions have the ability to incorporate professional and social support, and provide instructions to change behavior and problems in informal caregivers of people living with dementia [10] without requiring face-to-face delivery. Recent studies also indicated that web-based intervention programs can benefit the mental health of caregivers of adults living with chronic conditions, and particularly improve depression, stress and distress, and anxiety in caregivers [8,11]. Thus, web-based interventions aimed at supporting informal caregivers of people living with dementia have the potential to improve their psychological health; however, few interventions are widely implemented for this population [12,13]. A recent review assessing the role and effectiveness of web-based and app-based interventions in the self-management of dementia reported that few studies showed positive outcomes and were effective in improving self-management of people living with dementia [14]. Another meta-analysis that examined the effect of web-based interventions on the mental health outcomes of family caregivers of people living with dementia also found that most internet-based interventions were generally effective in reducing anxiety and depression in caregivers of people living with dementia [13]. However, little research has been done to *look under the hood* concerning the factors relating to behavioral change, such as behavior change theory, behavior change technique (BCT), and behavior change agent (BCA) that informs and shapes the interventions [15,16].

Behavior Change Theory, BCT, and BCA

In keeping with the definitions provided in Textbox 1, behavior change theories are abstract representations of interrelated concepts, definitions, and propositions that explain behavior change [17]. BCTs are observable, replicable, irreducible, and active ingredients within the intervention designed to change behavior [18]. A BCA is a putative mechanism or process that is measurable and modifiable and is hypothesized to play a causal role in producing behavior change [19]. To date, none of the prior reviews have systematically examined the use of the BCTs and BCAs in web-based interventions for informal caregivers of people living with dementia [12]. More specifically, the extent to which behavior change components are involved in the web-based interventions is still unclear [16]. Moreover, to the best of our knowledge, there is no report to date that explores the relationship of BCTs, BCAs, and retention. Failing to retain a sufficient number of participants in behavior interventions may not only lead to uncertainty about intervention effectiveness and pose a threat to the external validity of the results, but may also be associated with implementation challenges such as increased burden and low engagement [20]. Being able to identify and specify the behavioral active components of web-based interventions and cross-reference them with retention in the studies will provide a better mechanistic understanding of web-based interventions and allow future studies to be replicated more successfully in terms of retention across different settings and populations.

Textbox 1. Definitions for behavior change theory, behavior change technique (BCT), and behavior change agent (BCA).

Definitions

1. Behavior change theory: an abstract representation of an interrelated concept or theory explaining behavior change in the intervention (eg, Stress and Coping model, Cognitive Behavior Therapy Model, Adaption-Coping model, or Transition Theory).
2. BCT: an active, observable, replicable, and intricated component in the intervention which induces the behavior change (eg, goal planning, feedback on behavior, or demonstration of behavior).
3. BCA: a putative target or a mediator variable in the mechanism of behavior change (eg, self-efficacy, caregiver burden, or caregiver stress).

Considering the gaps in behavioral change mechanisms in web-based interventions for informal caregivers of people living with dementia and the increasing number of web-based interventions, it is timely to conduct a scoping review to explore and analyze the emerging literature [21]. This study reviewed the literature to map BCTs, theories, and agents deployed in the web-based interventions for informal caregivers of people living with dementia. Furthermore, there is an emerging consensus that in addition to intervention characteristics, retention could also be shaped by participant characteristics [22-25]. For example, in 2020, Ashford et al [22] identified that sociodemographic variables such as race and education level were associated with decreased task completion and enrollment in web-based interventions for older adults. Another study by Teles et al [25] describing the access and retention in psychosocial interventions for informal caregivers of people living with dementia suggested that caregiver education, their perceived mental health, and the number of hours spent in caregiving had a direct correlation with the retention or dropout rates of the study. As such, this study has three objectives: (1) to map behavioral theories, BCAs, and BCTs involved in web-based interventions for informal caregivers of people living with dementia; (2) to examine the relationship between behavioral change elements and retention in the studies; (3) to examine the relationship between participant characteristics and retention in the studies.

Methods

Overview

This scoping review followed guidelines from the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) [26], which was built upon prior scoping review frameworks of the Joanna Briggs Institute [27] and Arksey and O'Malley [28]. The PRISMA-ScR framework deleted 5 items (eg, risk of bias across studies, risk of bias within studies, and further analysis) from the original PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist, and it had the following main steps: (1) indicate whether a protocol and registration exist, (2) eligibility criteria, (3) information sources and search, (4) selection of sources of evidence, (5) charting data from the selected studies, and (6) synthesis of results [26].

Stage 1: Protocol and Registration

The protocol of this paper was modeled on our previous scoping review concerning behavior change factors and retention in dietary interventions for older adults [29]. Our protocol was drafted using the PRISMA Protocols [26]. The final protocol was registered prospectively with the Open Science Framework on February 18, 2022 (registered from the website [30]; registration DOI: 10.17605/OSF.IO/9M7K2).

Stage 2: Eligibility Criteria

Inclusion Criteria

Studies meeting the following inclusion criteria were included: (1) the intervention was aimed at informal caregivers (defined as a family member or friend providing unpaid care) of people living with dementia, (2) digital interventions delivered via the internet or apps, (3) the article considered a specific intervention and provided a description of it, (4) experimental design including quasi-experimental studies (ie, nonequivalent control with pretest-posttest design, nonequivalent control with posttest only, one group pre-post, and time series designs) and randomized controlled trials, (5) feasibility study, (6) published from January 1998 to March 2022, and (7) published in English.

Exclusion Criteria

Studies meeting the following exclusion criteria were excluded: (1) studies that focused on people living with early-onset dementia; (2) the intervention was solely delivered by telephone or was telemedicine based; (3) the interventions solely used Skype or other means of web-based calling; (4) the intervention had a large face-to-face component; (5) the results or outcomes of the intervention were not reported; (6) the intervention was focused on the person with dementia; (7) the study was not published in a peer-reviewed journal; (8) basic science articles (eg, animal studies, neuroanatomy, neuroimaging, anatomy, physiology, bacteriology, pathology, or biochemistry) fundamental to the study of medicine; (9) pertained to caregivers aged ≤18 years (per the definition of adults according to the National Institutes of Health); (10) focused on delirium, developmental disorders, or other; (11) letters to the editor, editorials, essays, or other op-ed pieces; (11) gray literature and review articles; (12) other (case study, proposed studies, or study protocol).

We excluded telephone-based support and extensive face-to-face interventions from our study as we intended to focus on digital technologies that could be used by caregivers without professional input.

Stage 3: Information Sources and Search

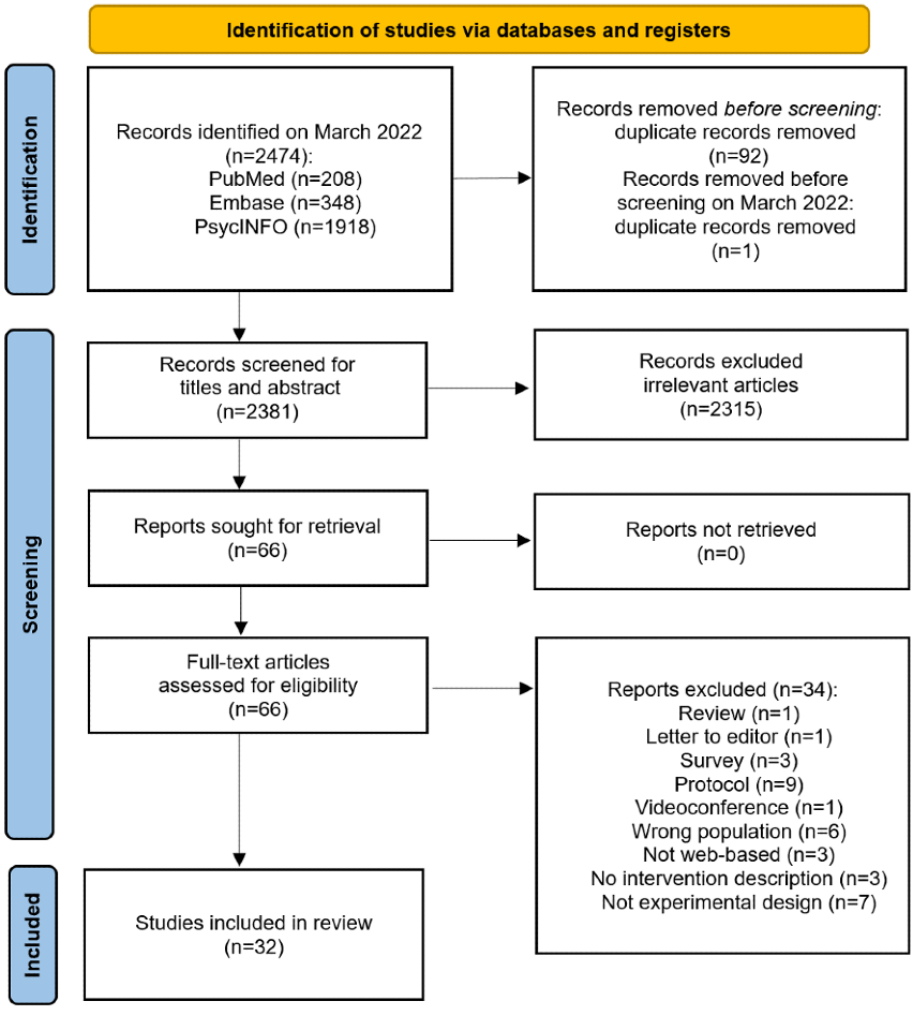
A systemic literature search was conducted in 3 databases to identify all relevant literature: PubMed, PsycINFO, and EMBASE. Keywords and Medical Subject Headings (MeSH) terms were used regarding the concepts of mobile health, telehealth, web-based, web, dementia, caregiver. The following were specific keywords used in the searching strategies:

(*caregiver* OR *caregivers* OR *carer* OR *carers* OR *Caregivers* [MeSH] OR *carepartner* OR *care partners*) AND (*Dementia* [MeSH] OR *dementia* OR *dementias*) AND (*"Internet-BasedIntervention* [MeSH] OR *online* OR *web-based* OR *internet* OR *on-line* OR *electronic* OR *MobileApplications* [MeSH] OR *mobile application* OR *mobile applications* OR *mobile app* OR *mobile apps* OR *tablet* OR *iPad*) AND (*support* OR *supportive* OR *Social Support* [MeSH] OR *Self-Help Groups* [MeSH]). Detailed search strategies are provided in Multimedia Appendix 1. Reference checking was performed to include potentially relevant studies. The research period was from January 1999 to March 2022, considering that the internet was widely used by the general public in 1999.

Stage 4: Selection of Sources of Evidence

All citations were uploaded to Rayyan [31], a web-based research tool that helps researchers to collaborate in systemic reviews and other knowledge synthesis projects. Duplicates were removed, and 2 reviewers (K-CW and YS) subsequently screened all the articles by title, abstract, and full text. The reviewers also reviewed each other's results before proceeding to the next step to avoid screening bias. When a disagreement occurred, they discussed the eligibility of the article regarding the research goal and the inclusion and exclusion criteria until a consensus was reached. In addition, a third reviewer (OZ) was involved to arbitrage disagreements between the 2 reviewers. The detailed screening process is illustrated in Figure 1.

Figure 1. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) flow diagram showing the study selection process.



Stage 5: Charting the Data

After reviewing all eligible studies, 2 reviewers (K-CW and YS) independently coded the literature using the BCTv1 taxonomy [18]. A BCTv1 taxonomy is a taxonomy methodology that comprises 93 individual BCTs grouped in 16 hierarchical clusters (eg, goals and planning, feedback and monitoring, social support, and shaping knowledge) that can specify the active ingredients in interventions. A data-reporting table was generated to guide the data abstraction process and display a summary of these study features: citation, study design, study location, sample characteristic, intervention characteristic, behavior change theory, retention rate, BCT, BCA, and outcome measures.

Stage 6: Synthesis of Results

We reported results in the following order: (1) sample characteristics; (2) outcomes; (3) theory; (4) BCT; (5) BCA; (6) retention; and (7) relationship among sample characteristics, theory, BCT, BCA, and retention. Sample characteristics included study design, location, caregiver characteristics (type, mean age, and male proportion) and composition. Outcome

measures were primary outcomes specified in the objectives and were measured before and after the intervention [32]. Behavior change theories were identified if they were explicitly referenced as the guiding theory or framework for an intervention. When a construct used in the intervention aligned with the behavior change theory or was mentioned in the intervention as a mediator, and was measured before and after the intervention, it would be considered a BCA. To further differentiate BCAs, we categorized each BCA into 3 BCA domains according to the methodology introduced by the Science of Behavior Change [33]. The retention rate was calculated as the percentage of participants who completed the study procedures as prescribed in the protocol. In this study, retention rates were either explicitly stated by the researchers or calculated from flow charts or comparable sources. The relationships between variables and retention rates were assessed using the Spearman rank correlation coefficient because of the observed monotonic but nonlinear trends between the variables, with $P<.05$ indicating statistical significance.

Results

Study Selection

The literature review from the PubMed, PsycINFO, and EMBASE databases yielded 2474 results, and 2381 articles were left after duplicates were removed. After abstract and title screening, 2315 irrelevant articles were removed, and 66 studies were retrieved for the full article review. In total, 32 articles were included in the final list after full article eligibility criteria were applied (see Figure 1 for the PRISMA-ScR flow diagram).

Sample Characteristics

Among the 32 articles, 16 (50%) were conducted in the United States, 10 (31%) were conducted in a European country, 4 (13%) in Canada, 1 (3%) in India, and 1 (3%) in Mexico and South America. In total, 17 studies used a randomized controlled trial in the study design, 8 studies were pilot studies, 4 were feasibility studies, 6 were mixed methods studies, and 3 studies were quasi-experimental. In total, 29 out of the 32 studies included a pretest or posttest design. The reported mean participant age ranged from 44 to 76 years, with a median of 62.4 (IQR 9.52) years and a mean age of 61.48 (SD 7.35) years. The sample sizes ranged from 10 to 486 participants with a median of 63 (IQR 109) participants. The average of reported male proportion was 24.93% (SD 11%; range 0%-53.6%). The mean of the reported spousal and adult children caregiver proportion were 51.24% (SD 25%) and 44.81% (SD 22%), respectively. The main characteristics of the included studies are presented in Multimedia Appendix 2 [34-65].

Interventions

Almost 80% (25/32) of the interventions were web-based interventions only, while the remaining 20% (7/32) integrated web-based intervention with other telehealth modalities (eg, telephone, virtual reality, email contact, and video conferences); 47% (15/32) of the studies did not include interventionists or facilitators when delivering their interventions. Only 10 studies reported their breakdown by sessions, which ranged from 4 to 8 modules. In addition, 81% (26/32) of the included studies reported the duration of their interventions. Table 1 presents the length of the intervention from the reported studies. The intervention durations varied from 2 weeks to 12 months with a median of 90 (IQR 138) days.

Table 1. Intervention period of reported studies (n=32).

Intervention duration	Studies, n (%)
12 weeks	7 (22)
24 weeks	6 (19)
N/A[a]	6 (19)
4 weeks	4 (13)
8 weeks	2 (6)
16 weeks	2 (6)
48 weeks	1 (3)
6 weeks	1 (3)
2 weeks	1 (3)
26 weeks	1 (3)
3 weeks	1 (3)

[a]N/A: not applicable.

Outcomes

There were at least 18 outcomes measured in different studies. The most common outcome type was the health indicator (25/32, 78%) of informal caregivers, including caregiver burden, stress, depression, pain, and loneliness. The second most measured outcomes were perceived competence (11/32, 34%). Other common outcomes found in the studies were problematic behaviors of care recipients (7/32, 22%), self-efficacy (7/32, 22%), perceived social support (6/32, 19%), quality of life (5/32, 16%), caregiving knowledge (4/32, 13%), quality of the relationship with care recipients (4/32, 13%), perceived health (4/32, 13%), and intervention usability and feasibility (4/32,

13%). Some peculiar outcomes include eHealth literacy, heart rate variability, costing, and cost-effectiveness.

Theories

In total, 53% (17/32) of the studies explicitly mentioned using behavior change theories or models to guide their interventions. Of these, only 4 studies mentioned more than one model. As Table 2 shows, the most used theory was the Stress and coping theory (6/17, 35%) and the CBT (6/17, 35%). Other theories or models were psychoeducational intervention (5/17, 29%), Transition theory (2/17, 12%), Trigger behavior response (1/17, 6%), Adaption-coping model (1/17, 6%), Social cognitive theory (1/17, 6%), Stress process model (1/17, 6%), and Communities of practice theory (1/17, 6%).

Table 2. Theories specified in guiding the interventions (n=32). An intervention could be guided by more than one theory.

Theory	Studies, n (%)
N/A[a]	15 (88)
Stress and coping model	6 (35)
Cognitive behavioral therapy	6 (35)
Psychoeducation	5 (29)
Transition theory	2 (12)
Trigger behavior response	1 (6)
Stress process model	1 (6)
Adaption-coping model	1 (6)
Social cognitive theory	1 (6)
Communities of practice theory	1 (6)

[a]N/A: not applicable.

BCAs and BCA Domains

We found BCAs in 81% (26/32) of the studies. The most common BCA was caregiver burden or strain (14/26, 54%). Other common BCAs included self-efficacy or confidence (10/26, 38%), caregiver stress or distress (10/26, 38%), social support (8/26, 31%), and caregiving competence or skill mastery (8/26, 31%). In addition, using the approach adopted from Nielsen et al [19] and the Science of Behavior Change taxonomy, which clusters BCAs into 3 major groups (interpersonal, stress reactivity, and self-regulation), we found that *stress reactivity* (20/26, 77%) was the most common BCA featured in 16 studies. The second commonly used BCA was *self-regulation* (18/26, 69%), and the least common was *interpersonal* (11/26, 42%).

Behavior Change Techniques

All 32 articles included at least one BCT, and 97% (31/32) of the studies included at least 2 BCTs in their intervention. The total number of individual BCT included in the studies ranged from 1 to 14, with a median of 5 (IQR 4) techniques. The total number of BCTv1 clusters ranged from 1 to 9 with a median of 4.5 (IQR 3) clusters. The individual BCT in each study is listed in Multimedia Appendix 3, and the frequency distribution of BCTv1 clusters specified in each study is presented in Table 3. The most frequently deployed BCTv1 cluster was *shaping knowledge* (27/32, 84%). The other common clusters include *social support* (19/32, 59%), *comparison of outcomes* (19/32, 59%), *comparison of behavior* (18/32, 56%), and *goals and planning* (17/32, 53%).

Table 3. Behavior change techniques (BCTs) taxonomy specified in guiding the interventions. One BCT cluster might appear in multiple studies (n=32).

BCT cluster	Studies, n (%)
1. Goals and planning	17 (53)
2. Feedback and monitoring	13 (41)
3. Social support	19 (59)
4. Shaping knowledge	27 (84)
5. Natural consequences	3 (9)
6. Comparison of behaviors	18 (56)
7. Associations	4 (13)
8. Repetition and substitution	4 (13)
9. Comparison of outcomes	19 (59)
10. Reward and threat	1 (3)
11. Regulation	1 (3)
12. Antecedents	10 (31)
13. Identity	10 (31)
14. Scheduled consequences	0 (0)
15. Self-belief	4 (13)
16. Covert learning	0 (0)

Retention

Retention rates were extracted from 91% (29/32) of the studies. The range of the retention rates from included studies varied from 32.6% to 97.4%, with an average of 70.44% (SD 17%) and a median of 74.6% (IQR 15%). Considering 80% as the third quantile of the retention rate in the included studies, we defined any study with an 80% retention or above as a high retention rate study. Only 28% (9/32) [34,41,44,48,49,52,53,55,63] of the interventions were high retention rate studies.

Retention With BCA, BCT, and Sample Characteristics

As shown in Figure 2, when examining high or low retention studies by 3 BCA domains (stress reactivity, self-regulation, and interpersonal), we found that stress reactivity was more common in low retention studies (n=16) while self-regulation (n=8) and interpersonal (n=6) were more common in high retention studies.

Figure 3 shows the relationships between retention rate and BCTv1 clusters or BCA domains. The Spearman coefficient suggested no significant relationship between the retention rates and the BCA domains (r=0.1; P=.60). However, there was a significant and negative relationship between the retention rate and the number of BCTv1 clusters (r=−0.47; P=.01).

Figure 4 presents the relationships of retention rate to age, gender, and spouse or adult children proportion. According to the Spearman correlation coefficient, no significant differences were found between the retention rate and informal caregiver's age (r=−0.03; P=.90) and gender (r=0; P=.99). However, we found that the proportion of spousal caregivers was significantly and negatively correlated with the retention rate (r=−0.45; P=.02), whereas the proportion of the adult children caregivers was significantly and positively correlated with the retention rate (r=0.5; P=.03) in the studies.

Figure 2. Behavior change agent (BCA) domains used in high vs low retention studies.

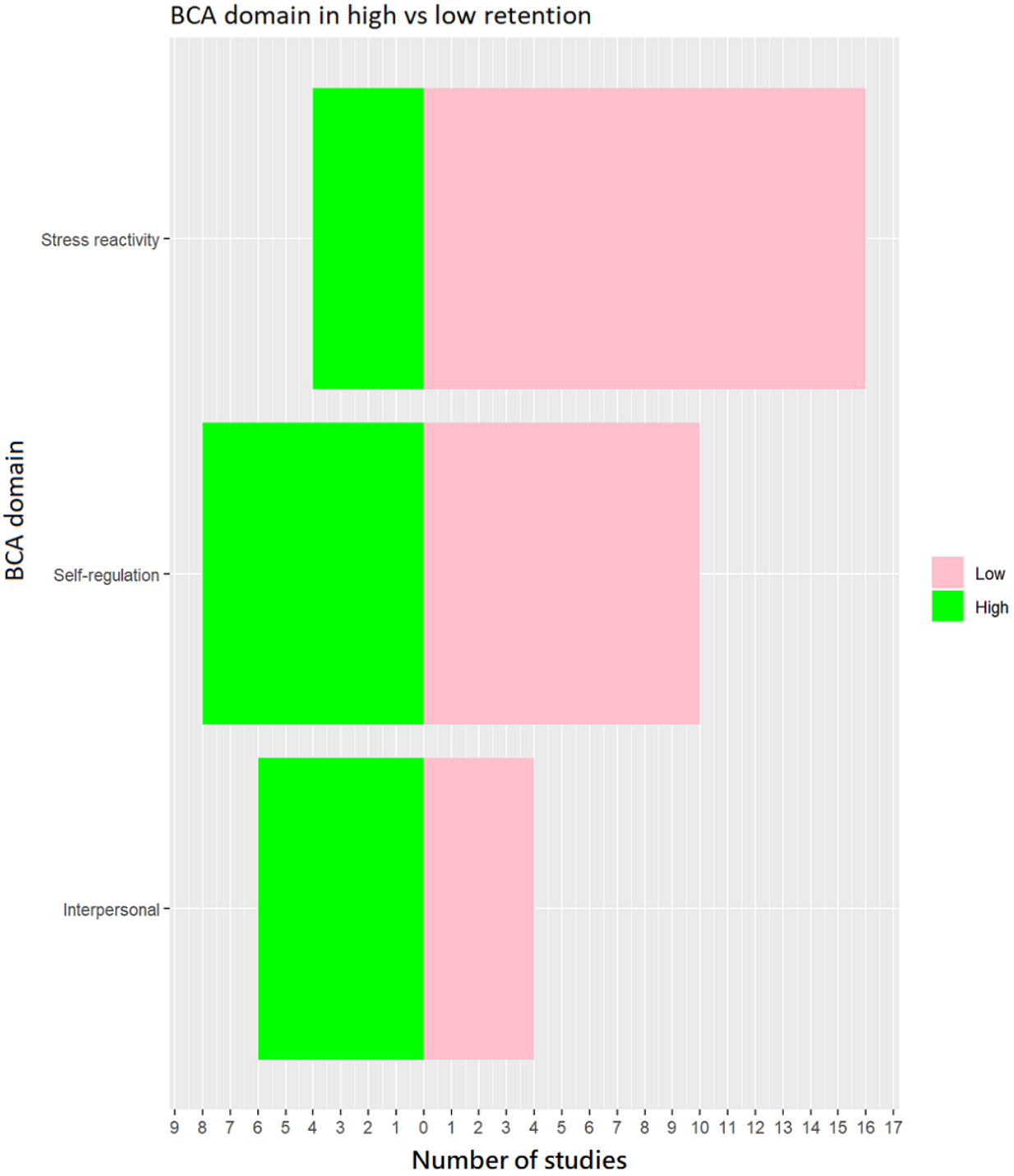


Figure 3. Relationships of retention rate to behavior change technique (BCT) clusters and behavior change agent (BCA) domains.

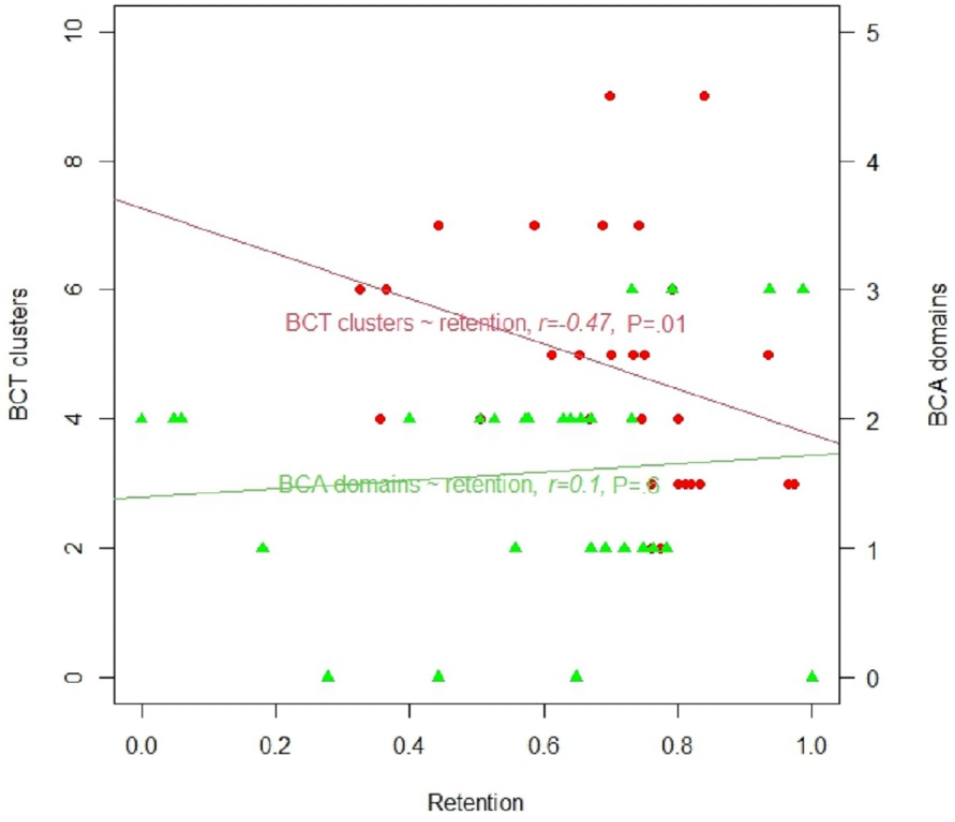
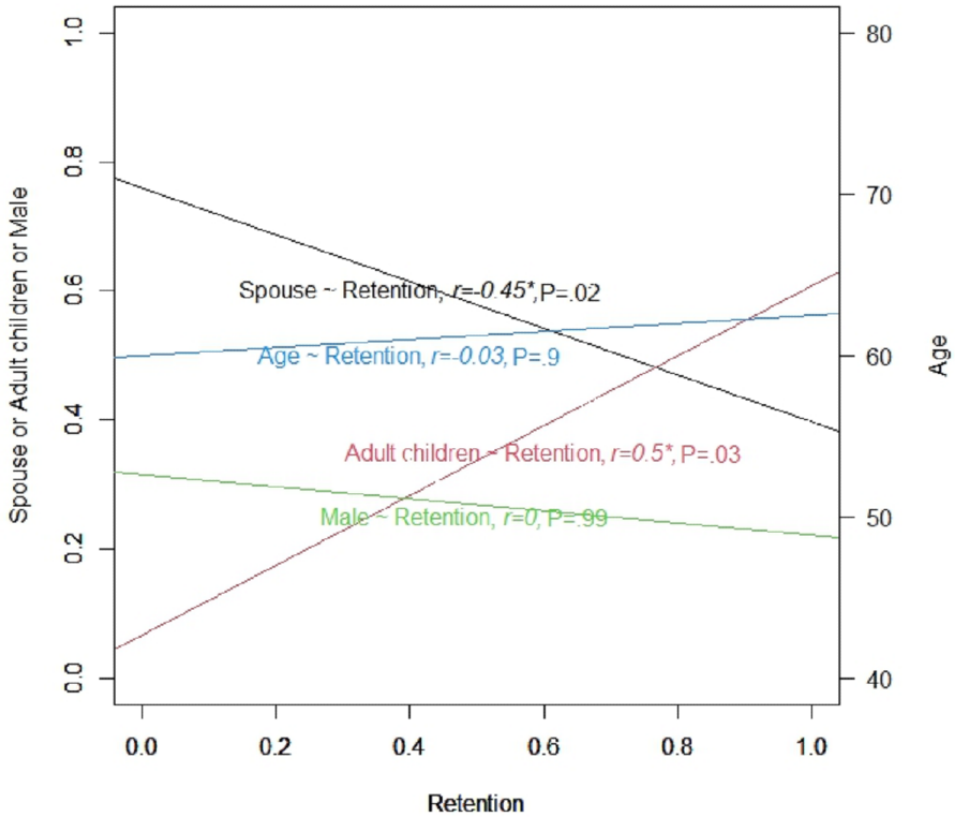


Figure 4. Relationships of retention rate to age, gender, spouse, and adult children proportion. *: *P*<.05.



Discussion

Principal Findings

The goals of this scoping review were to describe the level of evidence of behavioral factors (theory, BCA, and BCT) in web-based interventions for the informal caregivers of people living with dementia and to examine the relationship between sample characteristics, behavioral change factors, and study retention rates. We have 3 major findings. First, only about a half of the studies described their theoretical framework, but BCTs and BCA were more common. Second, we found that average retention rate has been around 70%, suggesting that it is a challenge for most web-based interventions in this population to retain participants. Third, the number of BCTv1 clusters and proportion of spousal caregivers are significantly and negatively correlated with retention rates while the involvement of adult children caregivers is significantly and positively correlated with the retention rate in the studies.

Comparison With Prior Work

The first finding indicates that about half of the web-based interventions lacked theory to guide their intervention. This finding is congruent with the results of 3 systematic reviews of health interventions in people with chronic conditions. In 2017, a review of exercise interventions in people living with dementia found 33% included studies used behavior change theories [66]. A review in 2019 of interventions targeting people with chronic neurological conditions and their caregivers found 59% of the studies did not mention theory, and only 22% were explicitly theory-based [67]. Another review of interventions among dementia caregivers revealed that only 37.5% of the studies used theory to inform interventions [68]. As mentioned by Walsh et al [15], there is a distinct lack of theoretical underpinnings in most dementia interventions [15], and interventions that make extensive use of theory may be able to have larger effects on behavior and improve the intervention sustainability than those that lack theory [15,69]. The explicit application of theory can help a study to better understand key aspects of the intervention, the participants, and the context, and offer a generalizable framework to inform the development of intervention as well as provide insights to possible causal mechanisms [15,70]. Therefore, we strongly advocate future research in the dementia caregiving context to include a theoretical basis in their intervention design.

In this study, we found that stress and coping theory, CBT, and psychoeducation are the most commonly adopted theories among web-based interventions for informal caregivers of people living with dementia. This finding partially aligns with several studies which specified CBT and psychoeducational approaches as the most effective and common theories in caregiver interventions which aimed to improve caregiving knowledge, well-being, and satisfaction as well as to reduce caregiver burden, anxiety, and depressive symptoms in caregivers of people living with dementia [5-7]. However, these reviews focused less on web-based interventions and may not provide the full picture of how theoretical underpinnings could guide interventions delivered on the web or in web-based settings. On the contrary, a meta-analysis released in 2021, which explored how web-based interventions improve mental health in home caregivers of people living with dementia, provided a good rationale for our finding in the *stress and coping theory*. The meta-analysis found that stress management program showed better outcomes in web-based interventions than other modified multicomponent integration programs [14]. Therefore, we suggest that future web-based intervention studies should retain the systemic integrity of the stress coping model when building their interventions.

In our review, 80% of the studies included BCA and all the studies included at least one BCT in their intervention, but the number of BCTs varied substantially. A recent review which assessed BCTs in clinical interventions confirmed their effectiveness in retention context [16] that is in sync with our results. Furthermore, both studies described a range BCTs in terms of numbers guiding their intervention design. For example, Fakolade et al [67] found that, “across 27 studies, two to 17 BCTs (mean 6.8, SD 4.02) were used.” We found the top three frequently deployed BCTv1 clusters were *shaping knowledge*, *social support*, and *comparison of outcomes*. These results are similar to a systematic review of internet-based interventions for caregivers of older adults [71]. This systematic review found the most frequently used BCTs included in efficacious interventions were provision of social support and the combination of instructions to guide behavior change and barrier identification. Another systematic review which mapped behavioral factors in health interventions for people with chronic neurological conditions and their caregivers shared analogous views that *shaping knowledge* and *comparison of outcomes* are 2 of the most common implemented BCT clusters [67]. To date, no reviews have mapped behavioral change factors with web-based interventions in dementia research. However, 2 systematic reviews that evaluated web-based self-management programs for parental caregivers to help children with diabetes and promote healthy eating in children both confirmed that *shaping knowledge* is a widely used and effective BCT cluster in web-based program interventions [72,73]. Future research in web-based interventions for dementia caregivers should consider retaining *shaping knowledge* while developing their interventions. Unfortunately, we did not find reviews which targeted interventions for people living with dementia and their caregivers stipulating BCA in their study. Considering increasing calls from the National Institutes of Health to emphasize on mechanisms of change [19], BCA should be explicitly specified and evaluated in future intervention development.

Our second major finding is that retention is still a challenge in most web-based interventions for caregivers of people living with dementia. We calculated an average of 70.44% (SD 17%) and a median of 74.6% (IQR 15%) retention rate in 32 studies. This result is similar to a cross-sectional study of retention of dementia caregivers in psychosocial interventions which reported high dropout rates (more than 50% for most intervention) in most psychosocial interventions [25]. Nevertheless, compared with the study by Teles et al [25], it seems the web-based interventions increased retention among this population. Our result about low retention rate contrasted with a systematic review which examines the effectiveness of

mobile and web-based health apps that support self-management and transition in young people with chronic physical health illnesses. The review reported an average 93% retention across 68 studies. The reasons that this systematic review observed higher retention compared with our study might result from the differences in our target populations (young people with chronic disease versus older adult dementia caregivers). Older adult participants may face more barriers in web-based interventions compared with younger generations owing to age-related changes (eg, changes in vision, hearing, and motor functions) [74]. Therefore, we might expect a lower retention in older participants in web-based interventions, especially in self-guided web-based interventions [75,76]. Future web-based interventions for informal caregivers of people living with dementia should consider the needs of this population and incorporate them in the development of the interventions.

The third major finding in our study is that retention rate was significantly associated with certain behavioral factors and sample characteristic. We identified a reduction in retention with an increase of BCTv1 clusters employed in interventions. The finding that the numbers of BCTv1 clusters is negatively associated with retention is at odds with the systematic review conducted by Duncan et al [16], which advocates for the application of BCTs to improve retention. Another meta-analysis which investigated how incorporating BCTs in internet-based programs could reduce smoking cessation in the public also presents different evidence from what was found in the study [77]. The meta-analysis reported that the number of BCTs in the long term was not significantly associated with treatment effectiveness (odds ratio 1.02, 95% CI 0.99-1.05; *P*=.16). However, the meta-analysis did not correlate the BCT numbers and BCTv1 clusters with retention rates. To conclude, our study raises several important issues for future research. For example, it would be important to examine the extent to which BCT potentially affect retention. If such an effect confirmed, factors facilitating retention rate such as particular BCTv1 clusters, the overall number of BCTv1 clusters or other features is another important consideration. Future studies should work on better elucidating the mechanisms of behavioral change, and explore how behavioral factors (theories, BCTs, and BCAs) affect the effectiveness and retention of the intervention, as well as provide more details on how BCTs should be leveraged.

As for sample characteristics, we found that the increase of spousal caregivers or the decrease of adult children caregivers in the study are significantly correlated with lower retentions. There were several studies that considered sample characteristics as predictors of retention. For example, a meta-analysis which identified predictors of treatment dropout in self-guided web-based interventions for depression found that being male, younger age, attained lower education level, and with comorbid anxiety symptoms were all predictors for a high dropout rate [75]. Some other studies also reported being young [23,24], less educated [22-24], race of people of color [22,23], or person with a lower socioeconomic status [24] were factors associated with lower retention rate in the intervention. A study of the access and retention of informal dementia caregivers in psychosocial interventions also reported significant associations between retention rate and behavioral factors of caregivers such as the number of hours spent in caregiving, and informational barriers [25]. However, we did not identify any study that associated retention rate with the ratio of spousal caregivers or adult children caregivers.

Limitations

Several limitations should be considered in this scoping review. One limitation was a lack of clarity in some interventions and theories. In some of the studies, interventions and the theoretical framework were not fully specified. Poorly justified intervention design and lack of detail descriptions of intervention could lead to challenges for researchers to abstract the BCTs or affect the analysis of behavioral factors in each intervention. Another limitation of this scoping review is the lack of other demographics information such as ethnicity, socioeconomic status, and education levels of caregivers. These are not available in many included studies but they might affect findings about adult children and spouse relationship with retention. A final limitation is the inconsistency in the intervention duration, which is the length of each web-based intervention. The length of the interventions ranged from 3 weeks to 12 months, which might ultimately impact the retention rate of the study. It is our recommendation for future studies to consider the length of the interventions and their effects on the retention rates.

It is worth mentioning that the sample characteristics of our review might not be generalizable to the older informal caregivers of people living with dementia (caregivers aged ≥65 years), because the reported mean age of informal caregivers of people living with dementia in the studies reviewed was 8 years younger compared with the mean age provided by the American Association of Retired Persons (AARP) and Alzheimer's Association. According to the Alzheimer's Association and AARP, women account for approximately two-third of dementia caregivers, with an average age of 69.4 years [1,78], while the reported mean age of informal caregivers of people living with dementia in our study was 62.26 (SD 7.36) years. Possible reasons that the mean age of the studies our sample were younger than the actual dementia caregiver population reported by the AARP and Alzheimer's Association are that web-based interventions may create specific challenges for the older populations to participate in the study [79,80]. For example, limited access to high-speed internet and video chatting, owning older technology that induces hardware and software incompatibilities, unfamiliarity with new technologies and motivational barriers, or visual impairments that affect the comprehension of interventions are all possible factors that reduce the participation of older adult dementia caregivers. Although Alzheimer's Association announced that over half of the caregivers are providing assistance to a parent or an in-law with dementia [1], they did not specify the proportions of spousal caregivers and adult children caregivers across the informal caregiver population. Our study, however, found that the means of the reported spousal caregiver and adult children caregiver proportions in 32 studies were both around 50% (51% and 45% individually), with the reported spousal caregiver proportion being 6% more than the adult children caregiver proportion.

Conclusions and Future Implication

This is the first study to comprehensively describe behavioral change factors (theory, BCA, and BCT) and identify the active ingredients in web-based interventions for the informal caregivers of people living with dementia using the BCTv1 taxonomy. This is also the first study to map the relationship of retention rates with behavioral change factors and sample characteristics.

In our study, we found that almost half of the studies were not informed by behavior change theories. We also observed that web-based interventions for informal caregivers of people living with dementia usually face retention challenges. Furthermore, we found that spousal involvement and a higher number of BCTs were each associated with lower retention rates.

On the basis of these findings, we proposed 3 suggestions for future studies. In planning future web-based interventions for informal caregivers of people living with dementia, researchers should (1) report the theoretical basis and behavioral change factors informing their study design; (2) ensure to address the needs of informal caregivers and people living with dementia in intervention development; and (3) model mechanisms of behavioral change and further explore how behavioral factors (theories, BCTs, and BCAs) and sample characteristics affect the effectiveness and retention of the intervention, as well as provide more details on how BCTs are applied.

Acknowledgments

This study was supported through innovation grant program from the University of Washington Garvey Institute for Brain Health Solutions and the University of Washington de Tornyay Center for Healthy Aging.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Key literature search terms and search results.
[PDF File (Adobe PDF File), 15 KB - jmir_v24i7e38595_app1.pdf]

Multimedia Appendix 2
Characteristics of included studies.
[PDF File (Adobe PDF File), 238 KB - jmir_v24i7e38595_app2.pdf]

Multimedia Appendix 3
The behavioral change factors (theory, techniques, and agents) identified in each study.
[DOCX File , 41 KB - jmir_v24i7e38595_app3.docx]

References

1. Monica MM, Díaz-Santos M, Vossel K. Alzheimer's Association 2021 Facts and Figures Report. David Geffen School of Medicine. URL: https://www.eastonad.ucla.edu/images/E-Newsletter/Summer_2021/ucla-easton-center-summer-2021-e-newsletter-final.pdf [accessed 2022-03-01]
2. Allen AP, Curran EA, Duggan A, Cryan JF, Chorcoráin AN, Dinan TG, et al. A systematic review of the psychobiological burden of informal caregiving for patients with dementia: focus on cognitive and biological markers of chronic stress. Neurosci Biobehav Rev 2017 Feb;73:123-164. [doi: 10.1016/j.neubiorev.2016.12.006] [Medline: 27986469]
3. Cross AJ, Garip G, Sheffield D. The psychosocial impact of caregiving in dementia and quality of life: a systematic review and meta-synthesis of qualitative research. Psychol Health 2018 Nov 27;33(11):1321-1342. [doi: 10.1080/08870446.2018.1496250] [Medline: 30260239]
4. Mausbach BT, Romero-Moreno R, Bos T, von Känel R, Ziegler MG, Allison MA, et al. Engagement in pleasant leisure activities and blood pressure: a 5-year longitudinal study in alzheimer caregivers. Psychosom Med 2017;79(7):735-741. [doi: 10.1097/psy.0000000000000497]
5. Kwon O, Ahn HS, Kim HJ, Park K. Effectiveness of cognitive behavioral therapy for caregivers of people with dementia: a systematic review and meta-analysis. J Clin Neurol 2017 Oct;13(4):394-404 [FREE Full text] [doi: 10.3988/jcn.2017.13.4.394] [Medline: 29057632]
6. Sörensen S, Conwell Y. Issues in dementia caregiving: effects on mental and physical health, intervention strategies, and research needs. Am J Geriatric Psychiatry 2011 Jun;19(6):491-496. [doi: 10.1097/jgp.0b013e31821c0e6e]
7. Walter E, Pinquart M. How effective are dementia caregiver interventions? An updated comprehensive meta-analysis. Gerontologist 2020 Nov 23;60(8):609-619. [doi: 10.1093/geront/gnz118] [Medline: 33226434]
8. Kwok T, Au A, Wong B, Ip I, Mak V, Ho F. Effectiveness of online cognitive behavioral therapy on family caregivers of people with dementia. Clin Interv Aging 2014 Apr:631. [doi: 10.2147/cia.s56337]

9. Khan K, Hall CL, Davies EB, Hollis C, Glazebrook C. The effectiveness of web-based interventions delivered to children and young people with neurodevelopmental disorders: systematic review and meta-analysis. J Med Internet Res 2019 Nov 01;21(11):e13478 [FREE Full text] [doi: 10.2196/13478] [Medline: 31682573]
10. Marsch LA. Leveraging technology to enhance addiction treatment and recovery. J Addict Dis 2012;31(3):313-318 [FREE Full text] [doi: 10.1080/10550887.2012.694606] [Medline: 22873192]
11. Sherifali D, Ali MU, Ploeg J, Markle-Reid M, Valaitis R, Bartholomew A, et al. Impact of internet-based interventions on caregiver mental health: systematic review and meta-analysis. J Med Internet Res 2018 Jul 03;20(7):e10668 [FREE Full text] [doi: 10.2196/10668] [Medline: 29970358]
12. Pleasant M, Molinari V, Dobbs D, Meng H, Hyer K. Effectiveness of online dementia caregivers training programs: a systematic review. Geriatr Nurs 2020 Nov;41(6):921-935. [doi: 10.1016/j.gerinurse.2020.07.004] [Medline: 32703628]
13. Zhao Y, Feng H, Hu M, Hu H, Li H, Ning H, et al. Web-based interventions to improve mental health in home caregivers of people with dementia: meta-analysis. J Med Internet Res 2019 May 06;21(5):e13415 [FREE Full text] [doi: 10.2196/13415] [Medline: 31066680]
14. Lee AR, Gerritzen EV, McDermott O, Orrell M. Exploring the role of web-based interventions in the self-management of dementia: systematic review and narrative synthesis. J Med Internet Res 2021 Jul 26;23(7):e26551 [FREE Full text] [doi: 10.2196/26551] [Medline: 34309575]
15. Walsh K, Timmons S, Byrne S, Browne J, Mc Sharry J. Identifying behavior change techniques for inclusion in a complex intervention targeting antipsychotic prescribing to nursing home residents with dementia. Transl Behav Med 2021 Mar 16;11(2):470-483. [doi: 10.1093/tbm/ibaa053] [Medline: 32589204]
16. Duncan EM, Bennett T, Gillies K. Assessing effective interventions to improve trial retention: do they contain behaviour change techniques? Trials 2020 Feb 21;21(1):213 [FREE Full text] [doi: 10.1186/s13063-020-4151-4] [Medline: 32085717]
17. Glanz K, Rimer B, Viswanath K. Health Behavior and Health Education: Theory, Research, and Practice. Hoboken: John Wiley & Sons; 2008.
18. 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 20;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
19. Nielsen L, Riddle M, King JW, NIH Science of Behavior Change Implementation Team, Aklin WM, Chen W, et al. The NIH science of behavior change program: transforming the science through a focus on mechanisms of change. Behav Res Ther 2018 Feb;101:3-11 [FREE Full text] [doi: 10.1016/j.brat.2017.07.002] [Medline: 29110885]
20. Coday M, Boutin-Foster C, Goldman Sher T, Tennant J, Greaney ML, Saunders SD, et al. Strategies for retaining study participants in behavioral intervention trials: retention experiences of the NIH Behavior Change Consortium. Ann Behav Med 2005 Apr;29 Suppl(2):55-65. [doi: 10.1207/s15324796abm2902s_9] [Medline: 15921490]
21. Verdejo C, Tapia-Benavente L, Schuller-Martínez B, Vergara-Merino L, Vargas-Peirano M, Silva-Dreyer AM. What you need to know about scoping reviews. Medwave 2021 Mar 30;21(2):e8144. [doi: 10.5867/medwave.2021.02.8144] [Medline: 33914717]
22. Ashford MT, Eichenbaum J, Williams T, Camacho MR, Fockler J, Ulbricht A, et al. Effects of sex, race, ethnicity, and education on online aging research participation. Alzheimers Dement (N Y) 2020 May 26;6(1):e12028 [FREE Full text] [doi: 10.1002/trc2.12028] [Medline: 32478165]
23. Grill J, Kwon J, Teylan M, Pierce A, Vidoni ED, Burns JM, et al. Retention of Alzheimer disease research participants. Alzheimer Dis Assoc Disord 2019;33(4):299-306 [FREE Full text] [doi: 10.1097/WAD.0000000000000353] [Medline: 31567302]
24. Robinson L, Adair P, Coffey M, Harris R, Burnside G. Identifying the participant characteristics that predict recruitment and retention of participants to randomised controlled trials involving children: a systematic review. Trials 2016 Jun 22;17(1):294 [FREE Full text] [doi: 10.1186/s13063-016-1415-0] [Medline: 27334018]
25. Teles S, Ferreira A, Paúl C. Access and retention of informal dementia caregivers in psychosocial interventions: a cross-sectional study. Arch Gerontol Geriatr 2021 Mar;93:104289. [doi: 10.1016/j.archger.2020.104289] [Medline: 33181425]
26. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Sep 04;169(7):467. [doi: 10.7326/M18-0850]
27. Peters MD, 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]
28. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Social Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
29. Zaslavsky O, Su Y, Kim B, Roopsawang I, Wu K, Renn B. Behavior change factors and retention in dietary interventions for older adults: scoping review. Gerontologist 2021 Sep 03:133. [doi: 10.1093/geront/gnab133] [Medline: 34477843]
30. A scoping review of behavioral change factors and retention in online interventions for informal caregivers of persons living with dementia. Center for Open Science. URL: https://archive.org/details/osf-registrations-9m7k2-v1 [accessed 2022-06-30]
31. 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]

32. Smith P, Morrow R, Ross D. Outcome measures and case definition. In: Field Trials of Health Interventions: A Toolbox (3 edition). Oxford, England : Oxford University Press; 2015.
33. Method. Science of behavior change (SOBC). URL: https://scienceofbehaviorchange.org/method/ [accessed 2022-04-03]
34. Beauchamp N, Irvine A, Seeley J, Johnson B. Worksite-based internet multimedia program for family caregivers of persons with dementia. Gerontologist 2005 Dec;45(6):793-801. [doi: 10.1093/geront/45.6.793] [Medline: 16326661]
35. Boots LM, de Vugt ME, Smeets CM, Kempen GI, Verhey FR. Implementation of the blended care self-management program for caregivers of people with early-stage dementia (partner in balance): process evaluation of a randomized controlled trial. J Med Internet Res 2017 Dec 19;19(12):e423 [FREE Full text] [doi: 10.2196/jmir.7666] [Medline: 29258980]
36. Lewis ML, Hobday JV, Hepburn KW. Internet-based program for dementia caregivers. Am J Alzheimers Dis Other Demen 2010 Dec 04;25(8):674-679 [FREE Full text] [doi: 10.1177/1533317510385812] [Medline: 21131674]
37. Marziali E, Garcia LJ. Dementia caregivers' responses to 2 internet-based intervention programs. Am J Alzheimers Dis Other Demen 2011 Feb 31;26(1):36-43 [FREE Full text] [doi: 10.1177/1533317510387586] [Medline: 21282276]
38. Blom MM, Zarit SH, Groot Zwaaftink RB, Cuijpers P, Pot AM. Effectiveness of an internet intervention for family caregivers of people with dementia: results of a randomized controlled trial. PLoS One 2015 Feb 13;10(2):e0116622 [FREE Full text] [doi: 10.1371/journal.pone.0116622] [Medline: 25679228]
39. Kajiyama B, Thompson LW, Eto-Iwase T, Yamashita M, Di Mario J, Marian Tzuang Y, et al. Exploring the effectiveness of an internet-based program for reducing caregiver distress using the iCare Stress Management e-Training Program. Aging Ment Health 2013 Jul;17(5):544-554 [FREE Full text] [doi: 10.1080/13607863.2013.775641] [Medline: 23461355]
40. Pot AM, Blom MM, Willemse BM. Acceptability of a guided self-help internet intervention for family caregivers: mastery over dementia. Int Psychogeriatr 2015 Feb 04;27(8):1343-1354. [doi: 10.1017/s1041610215000034]
41. Duggleby W, Jovel Ruiz K, Ploeg J, McAiney C, Peacock S, Nekolaichuk C, et al. Mixed-methods single-arm repeated measures study evaluating the feasibility of a web-based intervention to support family carers of persons with dementia in long-term care facilities. Pilot Feasibility Stud 2018 Oct 31;4(1):165 [FREE Full text] [doi: 10.1186/s40814-018-0356-7] [Medline: 30410783]
42. Duggleby W, Ploeg J, McAiney C, Peacock S, Fisher K, Ghosh S, et al. Web-based intervention for family carers of persons with dementia and multiple chronic conditions (My Tools 4 Care): pragmatic randomized controlled trial. J Med Internet Res 2018 Jun 29;20(6):e10484 [FREE Full text] [doi: 10.2196/10484] [Medline: 29959111]
43. Hattink B, Meiland F, van der Roest H, Kevern P, Abiuso F, Bengtsson J, et al. Web-based STAR E-learning course increases empathy and understanding in dementia caregivers: results from a randomized controlled trial in the Netherlands and the United Kingdom. J Med Internet Res 2015 Oct 30;17(10):e241 [FREE Full text] [doi: 10.2196/jmir.4025] [Medline: 26519106]
44. Chiu T, Marziali E, Colantonio A, Carswell A, Gruneir M, Tang M, et al. Internet-based caregiver support for Chinese Canadians taking care of a family member with Alzheimer disease and related dementia. Can J Aging 2009 Dec 01;28(4):323-336. [doi: 10.1017/s0714980809990158]
45. Eisdorfer C, Czaja SJ, Loewenstein DA, Rubert MP, Argüelles S, Mitrani VB, et al. The effect of a family therapy and technology-based intervention on caregiver depression. Gerontologist 2003 Aug 01;43(4):521-531 [FREE Full text] [doi: 10.1093/geront/43.4.521] [Medline: 12937331]
46. Finkel S, Czaja SJ, Schulz R, Martinovich Z, Harris C, Pezzuto D. E-care: a telecommunications technology intervention for family caregivers of dementia patients. Am J Geriatr Psychiatry 2007 May;15(5):443-448. [doi: 10.1097/JGP.0b013e3180437d87] [Medline: 17463195]
47. Griffiths PC, Whitney MK, Kovaleva M, Hepburn K. Development and implementation of tele-savvy for dementia caregivers: a department of veterans affairs clinical demonstration project. Gerontologist 2016 Feb 13;56(1):145-154. [doi: 10.1093/geront/gnv123] [Medline: 26566806]
48. Gustafson DH, Gustafson DH, Cody OJ, Chih M, Johnston DC, Asthana S. Pilot test of a computer-based system to help family caregivers of dementia patients. J Alzheimers Dis 2019 Jul 23;70(2):541-552. [doi: 10.3233/jad-190052]
49. Kales HC, Gitlin LN, Stanislawski B, Myra Kim H, Marx K, Turnwald M, et al. Effect of the WeCareAdvisor™ on family caregiver outcomes in dementia: a pilot randomized controlled trial. BMC Geriatr 2018 May 10;18(1):113 [FREE Full text] [doi: 10.1186/s12877-018-0801-8] [Medline: 29747583]
50. Lorig K, Thompson-Gallagher D, Traylor L, Ritter PL, Laurent DD, Plant K, et al. Building better caregivers: a pilot online support workshop for family caregivers of cognitively impaired adults. J Appl Gerontol 2010 Dec 15;31(3):423-437. [doi: 10.1177/0733464810389806]
51. O'Connor M, Arizmendi BJ, Kaszniak AW. Virtually supportive: a feasibility pilot study of an online support group for dementia caregivers in a 3D virtual environment. J Aging Stud 2014 Aug;30:87-93 [FREE Full text] [doi: 10.1016/j.jaging.2014.03.001] [Medline: 24984911]
52. Pagán-Ortiz ME, Cortés DE, Rudloff N, Weitzman P, Levkoff S. Use of an online community to provide support to caregivers of people with dementia. J Gerontol Soc Work 2014 Jul 29;57(6-7):694-709 [FREE Full text] [doi: 10.1080/01634372.2014.901998] [Medline: 24689359]

53. Huis In Het Veld JG, Willemse BM, van Asch IF, Groot Zwaaftink RB, Verkade P, Twisk JW, et al. Online self-management support for family caregivers dealing with behavior changes in relatives with dementia (part 2): randomized controlled trial. J Med Internet Res 2020 Feb 25;22(2):e13001 [FREE Full text] [doi: 10.2196/13001] [Medline: 32130142]
54. Núñez-Naveira L, Alonso-Búa B, de Labra C, Gregersen R, Maibom K, Mojs E, et al. UnderstAID, an ICT platform to help informal caregivers of people with dementia: a pilot randomized controlled study. Biomed Res Int 2016;2016:5726465 [FREE Full text] [doi: 10.1155/2016/5726465] [Medline: 28116300]
55. Wijma EM, Veerbeek MA, Prins M, Pot AM, Willemse BM. A virtual reality intervention to improve the understanding and empathy for people with dementia in informal caregivers: results of a pilot study. Aging Ment Health 2018 Sep 10;22(9):1115-1123. [doi: 10.1080/13607863.2017.1348470] [Medline: 28691861]
56. McKechnie V, Barker C, Stott J. The effectiveness of an internet support forum for carers of people with dementia: a pre-post cohort study. J Med Internet Res 2014 Feb 28;16(2):e68 [FREE Full text] [doi: 10.2196/jmir.3166] [Medline: 24583789]
57. Shaw CA, Williams KN, Lee RH, Coleman CK. Cost-effectiveness of a telehealth intervention for in-home dementia care support: findings from the FamTechCare clinical trial. Res Nurs Health 2021 Feb 19;44(1):60-70 [FREE Full text] [doi: 10.1002/nur.22076] [Medline: 33075157]
58. Williams KN, Shaw CA, Perkhounkova Y, Hein M, Coleman CK. Satisfaction, utilization, and feasibility of a telehealth intervention for in-home dementia care support: a mixed methods study. Dementia (London) 2021 Jul 09;20(5):1565-1585. [doi: 10.1177/1471301220957905] [Medline: 32902313]
59. Bruinsma J, Peetoom K, Boots L, Daemen M, Verhey F, Bakker C, et al. Tailoring the web-based 'Partner in Balance' intervention to support spouses of persons with frontotemporal dementia. Internet Interv 2021 Dec;26:100442 [FREE Full text] [doi: 10.1016/j.invent.2021.100442] [Medline: 34430221]
60. Baruah U, Varghese M, Loganathan S, Mehta KM, Gallagher-Thompson D, Zandi D, et al. Feasibility and preliminary effectiveness of an online training and support program for caregivers of people with dementia in India: a randomized controlled trial. Int J Geriatr Psychiatry 2021 Apr 08;36(4):606-617. [doi: 10.1002/gps.5502] [Medline: 33491811]
61. James TA, James D, Larkey LK. Heart-focused breathing and perceptions of burden in Alzheimer's caregivers: an online randomized controlled pilot study. Geriatr Nurs 2021 Mar;42(2):397-404. [doi: 10.1016/j.gerinurse.2021.02.006] [Medline: 33639543]
62. Fossey J, Charlesworth G, Fowler J, Frangou E, Pimm TJ, Dent J, et al. Online education and cognitive behavior therapy improve dementia caregivers' mental health: a randomized trial. J Am Med Dir Assoc 2021 Jul;22(7):1403-9.e1. [doi: 10.1016/j.jamda.2020.10.009] [Medline: 33288467]
63. Romero-Mas M, Ramon-Aribau A, de Souza DL, Cox A, Gómez-Zúñiga B. Improving the quality of life of family caregivers of people with Alzheimer's disease through virtual communities of practice: a quasiexperimental study. Int J Alzheimers Dis 2021;2021:8817491 [FREE Full text] [doi: 10.1155/2021/8817491] [Medline: 33884204]
64. Kajiyama B, Fernandez G, Carter EA, Humber MB, Thompson LW. Helping Hispanic dementia caregivers cope with stress using technology-based resources. Clin Gerontol 2018 Dec 13;41(3):209-216. [doi: 10.1080/07317115.2017.1377797] [Medline: 29236621]
65. Bass DM, Judge KS, Snow AL, Wilson NL, Morgan R, Looman WJ, et al. Caregiver outcomes of partners in dementia care: effect of a care coordination program for veterans with dementia and their family members and friends. J Am Geriatr Soc 2013 Aug 19;61(8):1377-1386. [doi: 10.1111/jgs.12362] [Medline: 23869899]
66. van der Wardt V, Hancox J, Gondek D, Logan P, Nair RD, Pollock K, et al. Adherence support strategies for exercise interventions in people with mild cognitive impairment and dementia: a systematic review. Prev Med Rep 2017 Sep;7:38-45 [FREE Full text] [doi: 10.1016/j.pmedr.2017.05.007] [Medline: 28593121]
67. Fakolade A, Walters AJ, Cameron J, Latimer-Cheung AE, Pilutti LA. Healthy together: a systematic review of theory and techniques used in health interventions for persons with chronic neurological conditions and their caregivers. Patient Educ Couns 2020 Apr;103(4):788-803. [doi: 10.1016/j.pec.2019.10.022] [Medline: 31767245]
68. Gitlin L, Marx K, Stanley I, Hodgson N. Translating evidence-based dementia caregiving interventions into practice: state-of-the-science and next steps. Gerontologist 2015 Apr;55(2):210-226 [FREE Full text] [doi: 10.1093/geront/gnu123] [Medline: 26035597]
69. 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]
70. Improved Clinical Effectiveness through Behavioural Research Group (ICEBeRG). Designing theoretically-informed implementation interventions. Implement Sci 2006 Feb 23;1:4 [FREE Full text] [doi: 10.1186/1748-5908-1-4] [Medline: 16722571]
71. Guay C, Auger C, Demers L, Mortenson WB, Miller WC, Gélinas-Bronsard D, et al. Components and outcomes of internet-based interventions for caregivers of older adults: systematic review. J Med Internet Res 2017 Sep 19;19(9):e313 [FREE Full text] [doi: 10.2196/jmir.7896] [Medline: 28928109]

72. Gomes AI, Pereira AI, Roberto MS, Boraska K, Barros L. Changing parental feeding practices through web-based interventions: a systematic review and meta-analysis. PLoS One 2021;16(4):e0250231 [FREE Full text] [doi: 10.1371/journal.pone.0250231] [Medline: 33909666]
73. van Vugt M, de Wit M, Cleijne WH, Snoek FJ. Use of behavioral change techniques in web-based self-management programs for type 2 diabetes patients: systematic review. J Med Internet Res 2013 Dec 13;15(12):e279 [FREE Full text] [doi: 10.2196/jmir.2800] [Medline: 24334230]
74. Rantakangas K, Halonen R. Special Needs of Elderly in Using Web-Based Services. Cham: Springer; 2020.
75. Karyotaki E, Kleiboer A, Smit F, Turner DT, Pastor AM, Andersson G, et al. Predictors of treatment dropout in self-guided web-based interventions for depression: an 'individual patient data' meta-analysis. Psychol Med 2015 Oct;45(13):2717-2726. [doi: 10.1017/S0033291715000665] [Medline: 25881626]
76. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
77. McCrabb S, Baker A, Attia J, Skelton E, Twyman L, Palazzi K, et al. Internet-based programs incorporating behavior change techniques are associated with increased smoking cessation in the general population: a systematic review and meta-analysis. Ann Behav Med 2019 Feb 01;53(2):180-195. [doi: 10.1093/abm/kay026] [Medline: 29750240]
78. Caregiving in the United States 2020. AARP. 2020. URL: https://doi.org/10.26419/ppi.00103.001 [accessed 2022-06-29]
79. Christie HL, Martin JL, Connor J, Tange HJ, Verhey FR, de Vugt ME, et al. eHealth interventions to support caregivers of people with dementia may be proven effective, but are they implementation-ready? Internet Interv 2019 Dec;18:100260 [FREE Full text] [doi: 10.1016/j.invent.2019.100260] [Medline: 31890613]
80. Remillard ML, Mazor KM, Cutrona SL, Gurwitz JH, Tjia J. Systematic review of the use of online questionnaires of older adults. J Am Geriatr Soc 2014 Apr 17;62(4):696-705 [FREE Full text] [doi: 10.1111/jgs.12747] [Medline: 24635138]

Abbreviations

AARP: American Association of Retired Persons
BCA: behavior change agent
BCT: behavior change technique
CBT: cognitive behavioral therapy
MeSH: Medical Subject Headings
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

Edited by A Mavragani; submitted 11.04.22; peer-reviewed by S Badawy, P Martinez; comments to author 16.05.22; revised version received 09.06.22; accepted 13.06.22; published 07.07.22.

Please cite as:
Wu KC, Su Y, Chu F, Chen AT, Zaslavsky O
Behavioral Change Factors and Retention in Web-Based Interventions for Informal Caregivers of People Living With Dementia: Scoping Review
J Med Internet Res 2022;24(7):e38595
URL: https://www.jmir.org/2022/7/e38595
doi:10.2196/38595
PMID:35797100

Review

Language Use in Conversational Agent–Based Health Communication: Systematic Review

Yi Shan[1], Prof Dr; Meng Ji[2], PhD; Wenxiu Xie[3], MPhil; Xiaobo Qian[4], BEng; Rongying Li[5], BM; Xiaomin Zhang[6], PhD; Tianyong Hao[4], Prof Dr

[1]School of Foreign Studies, Nantong University, Nantong, China
[2]School of Languages and Cultures, University of Sydney, Sydney, Australia
[3]Department of Computer Science, City University of Hong Kong, Hong Kong, China
[4]School of Computer Science, South China Normal University, Guangzhou, China
[5]School of Artificial Intelligence, South China Normal University, Guangzhou, China
[6]Department of Linguistics, Macquarie University, Sydney, Australia

Corresponding Author:
Yi Shan, Prof Dr
School of Foreign Studies
Nantong University
No. 9, Seyuan Rd.
Nantong, 226019
China
Phone: 86 15558121896
Email: victorsyhz@hotmail.com

Abstract

Background: Given the growing significance of conversational agents (CAs), researchers have conducted a plethora of relevant studies on various technology- and usability-oriented issues. However, few investigations focus on language use in CA-based health communication to examine its influence on the user perception of CAs and their role in delivering health care services.

Objective: This review aims to present the language use of CAs in health care to identify the achievements made and breakthroughs to be realized to inform researchers and more specifically CA designers.

Methods: This review was conducted by following the protocols of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 statement. We first designed the search strategy according to the research aim and then performed the keyword searches in PubMed and ProQuest databases for retrieving relevant publications (n=179). Subsequently, 3 researchers screened and reviewed the publications independently to select studies meeting the predefined selection criteria. Finally, we synthesized and analyzed the eligible articles (N=11) through thematic synthesis.

Results: Among the 11 included publications, 6 deal exclusively with the language use of the CAs studied, and the remaining 5 are only partly related to this topic. The language use of the CAs in these studies can be roughly classified into six themes: (1) personal pronouns, (2) responses to health and lifestyle prompts, (3) strategic wording and rich linguistic resources, (4) a 3-staged conversation framework, (5) human-like well-manipulated conversations, and (6) symbols and images coupled with phrases. These derived themes effectively engaged users in health communication. Meanwhile, we identified substantial room for improvement based on the inconsistent responses of some CAs and their inability to present large volumes of information on safety-critical health and lifestyle prompts.

Conclusions: This is the first systematic review of language use in CA-based health communication. The results and limitations identified in the 11 included papers can give fresh insights into the design and development, popularization, and research of CA applications. This review can provide practical implications for incorporating positive language use into the design of health CAs and improving their effective language output in health communication. In this way, upgraded CAs will be more capable of handling various health problems particularly in the context of nationwide and even worldwide public health crises.

(J Med Internet Res 2022;24(7):e37403) doi:10.2196/37403

KEYWORDS

systematic review; health communication; language use; conversational agent

Introduction

Background

Conversational agents (CAs) are intelligent computer programs empowered with natural language processing techniques that engage users in human-like conversations to provide an effective and a smart communication platform in a simulated environment, including text-based chatbots, voice-activated assistants, and embodied CAs [1,2]. They are designed to obtain specific information from users that is necessary to perform particular tasks and respond in a manner that is optimal to achieve these goals. Due to their ability to transform the health care system and enable individuals to comanage their health care effectively, CAs are increasingly used to deliver health care services [3]. The most popular health CAs include ELIZA [4], Casper [5], MedChat [5], PARRY [6], Watson Health [7], Endurance [7], OneRemission [8], Youper [9], Florence [10], Your.Md [11], AdaHealth [12], Sensely [13], and Buoy Health [14], among many others. CAs are being tested and adopted to provide and collect health-related information and provide treatment and counseling services [15]. In some cases, they are used to enhance the accessibility, efficiency, and personalization of service delivery and ensure relatively equal delivery of health care services worldwide through bridging the gaps between developing and developed countries [15,16].

Given the growing significance of CAs, researchers have conducted a plethora of relevant studies, varying from their suitability as health care partners to their designs including physical appearance, gender, and speech. [17-20]. These studies aimed to improve "humanness heuristics," affective states in users, and user perceptions of the CA personalities by tailoring CAs to the cultures and demographics of the users to continuously promote user engagement, adherence, and adoption [21-24].

Language plays a crucial role in improving user engagement because perceived impersonal closeness, intention to use, user satisfaction, establishment of trust, and user self-disclosure or self-concealment are closely associated with the task- and social-based interactivity, interaction, politeness, and information quality provided by CAs [2,19,21,25-32]. However, few studies focused on language use in CA-based health communication to examine its influence on the perceived usability of CAs and the perceived roles of CAs in delivering health care services [16]. Language considerably influences the joint construction of meaning between interlocutors and rapport establishment [23,33,34]. This is particularly true for human-machine communication. For example, when addressing users by their first names, CAs are perceived to display varying degrees of politeness and thoughtfulness determined by cultural limits and preferences [24]. It follows that intensive and extensive investigations into language use by CAs in different linguistic settings are crucial to scale up health care interventions delivered by CAs worldwide [16,35].

Language Use and Its Significance in CA Communication

The language use of an information source is likely to be crucial among various factors affecting the information seekers' judgments on the credibility and trustworthiness of the information providers [36-39]. In this review, language use, characterized by various linguistic aspects, is defined as varied verbal strategies and compliance-gaining techniques [40] that the CAs under scrutiny adopted to deliver health interventions. These strategies and techniques may involve various ways of wording, including an everyday style (eg, "heart attack") versus a technical style (eg, "myocardial infarction"), a tentative style (eg, "presumably similar") versus a nontentative style (eg, "similar"), a neutral style (eg, "methodological mistakes") versus an aggressive style (eg, "really dumb methodological mistakes"), an emotional style versus a nonemotional style, and an enthusiastic style versus a nonenthusiastic style. [36,41-44]. They may also include the use of personal references (eg, first-person and second-person pronouns), personal testimonials, specific conversational frameworks or prompts, and other verbal means of communication [45-47]. In short, the language use of the CAs under discussion in this review refers to their characteristic linguistic performances in health communication.

Language Expectancy Theory [48] and Communication Accommodation Theory [49] assert that acquiring knowledge when seeking web-based health information is determined not only by the information content but also by who is communicating the information and the manner and context of communication. Information seekers evaluate information providers positively if the latter's language use is in tune with their cultural values and situational norms and if they use language more favorably than expected in a situation [50]. The language use of information providers is regarded as a prominent clue to evaluate the characteristics of the providers, especially in web-based communication [51,52]. The information provider's language use is a cue for determining whether people perceive the information to be credible and whether the information provider is trustworthy [37,46].

Objective

The current review aimed to summarize the language use of CAs in health care to identify the achievements made and the breakthroughs to be made to inform researchers and more particularly CA designers and developers. This can help realize the high potential of CAs for improving individual well-being.

Methods

Study Design

The primary objective of the current review was to identify the language use of CAs in health care. This review was performed by following the protocols of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 statement [53]. We first designed a search strategy according to the research aim and then performed keyword searches in PubMed and ProQuest databases for retrieving relevant publications. Then, 3 researchers screened and reviewed the publications independently to select studies meeting the

predefined selection criteria. Finally, we synthesized and analyzed the eligible articles.

Search Strategy and Study Selection Criteria

This review focused on two aspects of the previous studies: CA applications in health care and language use. To retrieve a high number of relevant studies, we decided on using the keywords relating to language use for literature search, including "expression," "language," "language style," "language feature," "language characteristic," "language pattern," "linguistic style," "linguistic feature," and "linguistic characteristic." Based on these keywords and those concerned with CA applications in health care, we developed the following search strategy to identify studies wholly or partly investigating the language use of CAs: ((expression [Title/Abstract]) OR (language [Title/Abstract]) OR (language style [Title/Abstract]) OR (language feature [Title/Abstract]) OR (language characteristic [Title/Abstract]) OR (language pattern [Title/Abstract]) OR (linguistic style[Title/Abstract]) OR (linguistic feature [Title/Abstract]) OR (linguistic characteristic [Title/Abstract])) AND ((health* chatbot [Title/Abstract]) OR (health* conversational agent [Title/Abstract])). Drawing on this search strategy, we conducted keyword searches in 2 databases (PubMed and ProQuest) to retrieve published papers without restrictions regarding the year of publication on February 11, 2022.

We included both peer-reviewed and non–peer-reviewed journal publications because the aim of this review was to provide a comprehensive overview of the language use of CAs in health care and its corresponding implications for improvement in language use in CA communication to inform future research and CA designers. Textbox 1 shows the inclusion and exclusion criteria.

Textbox 1. Inclusion and exclusion criteria of the study.

Inclusion criteria

- Articles wholly or partly examining the language use of conversational agents (CAs) in health care were included.
- Articles on CAs that are equipped with languages other than English were included.

Exclusion criteria

- Publications that are not journal articles (eg, reports, editorials, dissertations, and news) were excluded.
- Articles that were not written in English were excluded.
- Articles that focus on the development of CAs and do not cover any design or setting of system-human linguistic interactions were excluded.
- Studies that examine the application of CAs in other fields than health care were excluded.

Article Selection and Data Extraction

We used Microsoft Excel (Microsoft Corporation) to manage the collected articles by listing the titles, abstracts, and article types for screening. First, 2 researchers (MJ and YS) screened the titles and abstracts of the candidate articles independently, filtering those articles that did not conform to the selection criteria. If the eligibility of some studies was unclear, we included them for further full-text review. Then, 2 researchers (YS and WX) reviewed the full texts of the remaining articles independently. Any disagreements were resolved through discussion and consultation with the third researcher (MJ).

To analyze and synthesize the language use of the health care CAs, the following information was extracted from eligible studies by YS: first author, year of publication, health care application, target population, study design, major findings, and limitations. Then, MJ reviewed and cross-checked the extracted data. Any discrepancies were resolved through a discussion with the entire research team.

Data Analysis and Synthesis

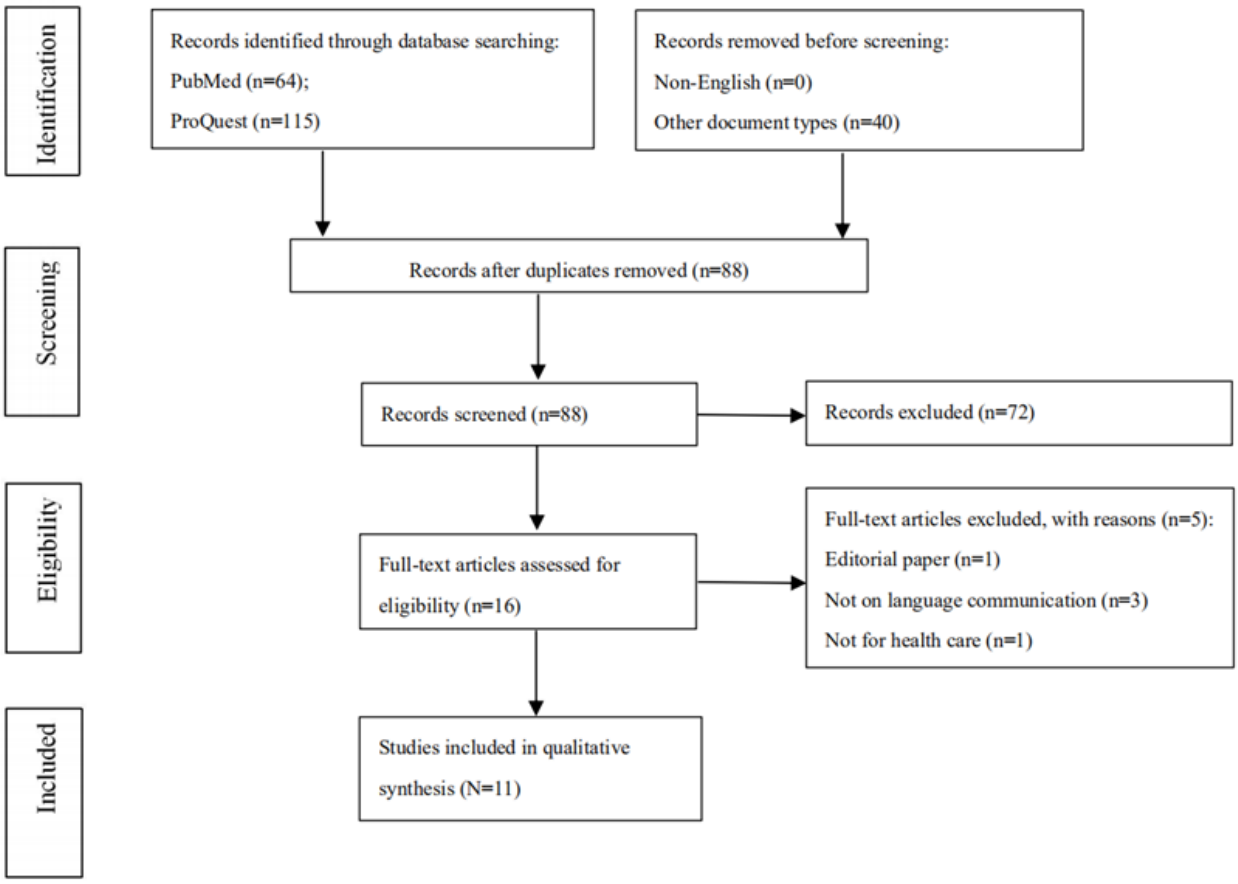
A meta-analysis was not feasible due to the expected variety of health care applications, target populations, study designs, results, and limitations. Therefore, we conducted thematic synthesis to summarize the data extracted from the included articles following 3 steps, namely "line-by-line" coding of the text, development of "descriptive themes," and generation of "analytical themes" [54]. YS first coded each line of the extracted text according to its meaning, then developed descriptive themes, and finally generated analytical themes using the derived descriptive themes [55]. MJ validated each assigned code, each derived descriptive theme, and each developed analytical theme independently. All the authors discussed and finalized the results of the thematic synthesis.

Results

Search Results

Using the search strategy, we identified 179 publications in the PubMed and ProQuest databases. From these retrieved publications, 40 were eliminated because they were not journal articles but were other types of publications (eg, commentaries, letters, news, and editorials); 51 were eliminated for being duplicates, and 72 for not meeting the selection criteria. After the full-text review, another 5 studies were excluded; 3 were not related to language communication, 1 was not about health care, and 1 was an editorial. As a result, 11 studies met the inclusion criteria and were eligible to be considered in this systematic review. Figure 1 shows the screening and selection process.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the selection of eligible studies.



Characteristics of Included Studies

Table 1 summarizes the information extracted from the 11 papers selected for synthesis and analysis. The major findings reported in the original studies that are directly related to the aim of the current review are included. We have included the limitations reported in the original studies and those based on our perspectives, if any. Based on this table, we present the qualitative synthesis and analysis in the Discussion section.

Table 1. Information extracted from the 11 selected studies.

Reference, first author, and year of publication	Health care application	Target population	Study design	Major findings	Limitations
[56]; Ollier; 2022	A public health CA[a] prototype	People in French and German lingua cultures	An internet-based experiment	The CA's choice of formal and informal forms of the second-person pronoun "You"— *Tu/Vous* (T/V) distinction—affected the evaluations of users of different ages, genders, and cultures to varying degrees.	Given that the study involved a complicated 4-way interaction between T/V distinction, language and culture, age, and gender, the sample size is not sufficiently large to ensure more generalizable findings. Therefore, the implications for CA designers are affected.
[57]; Kocaballi; 2020	Commonly available, general-purpose CAs on smartphones and smart speakers	Unspecified	Following a piloted script to present health- and lifestyle-related prompts to 8 CAs	The ratio of the CAs' appropriate responses decreased when safety-critical prompts were rephrased or when the agent used a voice-only interface. The appropriate responses included mostly directive content and empathy statements for the safety-critical prompts and a mix of informative and directive content for the lifestyle prompts.	Some response structures were derived from the patterns observed in the responses to a reasonably limited set of studied prompts, possibly not capturing additional or different structural elements of the CAs' responses based on a larger set of prompts. CAs failed to provide a larger amount of precoded information on some safety-critical prompts.
[58]; Boustani; 2021	An expressive, speech-enabled digital health agent to deliver an internet-based brief behavioral health intervention for alcohol use	51 alcohol users in the United States	Description of the CA design, acceptability, feasibility, and utility	The CA used a model of empathetic verbal and nonverbal behaviors to engage users, who had overwhelmingly positive experiences with the digital health agent, including engagement with the technology, acceptance, perceived utility, and intent to use the technology.	It is unclear whether the model of empathetic verbal and nonverbal behaviors the CA used to engage young and middle-aged adults successfully can equally engage the elderly or children in America and users in other countries, especially considering different cultural factors that may influence the perception of language.
[59]; Miner; 2016	68 phones from 7 producers	Investigators	A pilot study followed by a cross-sectional study	Some CAs replied to users' concerns with respectful language and referred them to helplines, emergency services, and nearby medical facilities, but some failed to do so.	Investigators used standardized phrases for health and interpersonal violence concerns, but people asking for help on their personal smartphones may use different phrases, which may influence the CAs' responses. The study only tested a limited number of CAs available in the United States and evaluated their responses to a limited number of health concerns, which may affect the generalizability of the findings.
[60]; Grové; 2020	A mental health and well-being chatbot named Ash	Young people aged 15-17 years and living in Australia	Interviews and a survey	The chatbot failed to identify and understand critical words and generate responses appropriate to critical words. Exemplary dialogs show the chatbot's respectful, empathetic, supportive, and encouraging language style.	The imbalanced numbers of male and female participants who interacted with the chatbot may influence the responses of the chatbot.
[61]; Ireland; 2015	Chatbots for people with Parkinson disease	People with Parkinson disease	A description of chatbots for people with Parkinson disease	The chatbots can engage with patients in random, human-like conversations.	The study failed to cite chatbot-patient conversations to illustrate the randomness and human-likeness of the conversations.
[62]; Frick; 2021	CAs	German participants	An internet-based questionnaire and a comparative study	Only an exemplary anamnesis with a CA shows the CA's polite, respectful, and encouraging language style.	The study failed to discuss the role of the CA's language style in soliciting disclosure of medical information from patients.

Reference, first author, and year of publication	Health care application	Target population	Study design	Major findings	Limitations
[63]; Cooper; 2018	A chatbot named Alex	Children on the autism spectrum	A description of a chatbot	The chatbot is able to engage with the user on a variety of topics using symbols and images.	The study aimed to describe a new chatbot and did not provide real-time exemplary conversations between the chatbot and real patients, making it difficult for us to understand the role of its language style in engaging patients.
[64]; Almushrraf; 2020	A motivational interviewing–based chatbot	Adult cigarette smokers	A single-arm prospective iterative design study	Due to the running head start technique that the chatbot used when engaging in conversations, 34.7% (42/121) of participants enjoyed the interaction with the chatbot. The chatbot finished the conversation after receiving the response to the exception case questions.	The running head start technique might not be appropriate or helpful for those who were already exhibiting change behavior. The lack of follow-on to the exception case questions or elsewhere in the conversation can frustrate subjects and possibly lead to negative unintended effects.
[65]; Ireland; 2016	An artificial CA named Harlie	People with neurological conditions such as Parkinson disease and dementia	A description of a chatbot	The chatbot is able to converse with the user on a variety of topics. It can engage patients with a random, human-like, well-manipulated conversation style to gain information about challenges patients encounter and play an educational and supportive role.	The study focused on the chatbot's role in performing different tasks without attaching importance to the function of its language style in engaging the patients.
[66]; Ireland; 2021	A trainee chatbot named Edna	5 genetic counselors and adults who had whole exome sequencing conducted for diagnosis of a genetic condition, either for themselves or their child	A description of a chatbot	The chatbot can engage users with a polite, respectful, and an encouraging language style. The chatbot can educate users through explaining genetic conditions and terminologies precoded into its language resources.	The chatbot cannot engage in conversations related to the impact of specific genetic conditions, emotive personal circumstances, or expert medical advice, which possibly influences its language style.

[a]CA: conversational agent.

Discussion

Principal Findings

In human-CA linguistic communication, language cues are particularly important because they perform a crucial function in promoting user engagement [2], but few studies examine significant sociolinguistic dimensions in CA design across different languages and cultures, and the impact of these dimensions on user perceptions of CAs and their effectiveness in delivering health care services [16]. In this review, 6 of the 11 included publications deal exclusively with the language use of the CAs studied, and the remaining 5 are partly related to this topic. We derived the following themes from the language use in the 11 included studies through thematic synthesis.

Personal Pronouns

Among the 6 studies exploring exclusively the language used by CAs, the most interesting and distinctive study analyzes the influence of the CA's use of formal and informal forms of the second-person pronoun "you"—*Tu/Vous* (T/V) distinction—across language contexts on user evaluations of digital health applications [56]. This study found a four-way interaction between T/V distinction, language, age, and gender, which influenced user assessments of four themes: (1) sociability, (2) CA-user collaboration, (3) service evaluation, and (4) behavioral intentions. Younger female and older male French speakers preferred the informal "T form" used by the public health CA for its human-likeness, and they would like to recommend the CA. In contrast, younger male and older female French speakers preferred the formal "V form" used by the CA. Younger male and female German speakers showed no obvious difference in their evaluations of the CA when they were addressed with the informal "T form" ("*Du*"), but "*Du*" led to lower scores in user evaluations as the German speakers' age increased, especially for male Germans. German speakers' user evaluation scores induced by the formal "V form" ("*Sie*") were relatively stable and not affected by gender, but they

increased slightly with age. The T/V distinction in French, German, Spanish, Chinese, Malaysian, and Korean, among many other distinctions of linguistic forms in various languages, indicates more or less formality, distance, or emotional detachment [67,68]. Such distinction encodes interactive meanings and shapes normative expectations such as politeness etiquette, the breach of which potentially results in perceived insult, membership of a different social class, and affiliation with another culture or grouping, leading to outcomes such as customer dissatisfaction [68-74]. CA developers need to consider this distinction and many other linguaculture-specific distinctions in the designing stage to enable CAs to choose appropriate forms for specific user groups, which facilitates user engagement in CA-based health communication [65,70].

Responses to Health and Lifestyle Prompts

A recent study analyzed the content appropriateness and presentation structures of CAs' responses to health and lifestyle prompts (questions and open-ended statements) [57]. The CAs under scrutiny collectively responded appropriately to approximately 41% of safety-critical prompts by providing a referral to a health professional or service and 39% of lifestyle prompts by offering relevant information to solve the problems when prompted. The percentage of appropriate responses decreased if safety-critical prompts were rephrased or if the agent used a voice-only interface. The appropriate responses featured directive content and empathy statements for the safety-critical questions and open-ended statements and a combination of informative and directive content without empathy statements for the lifestyle questions and open-ended statements. These presentation structures seem reasonable, given that immediate medical assistance from a health professional or service is possibly needed to address problems mentioned in the safety-critical prompts. The use of empathy aligns with the testified exploitation of empathy on sensitive topics, showing that empathy is an important defining determinant of an effective CA [66,75,76]. The CAs examined in this study also displayed some defects, including the same CA's inconsistent responses to the same prompt [57], which was also found in another study [40], and different answers from the same CA on different platforms. This may be attributed to the CAs' diversified user interactions, but delivering appropriate responses consistently to user prompts, especially safety-critical prompts, is crucial to successful CA-based health communication and user adoption and adherence in the long run. Another weakness was the CAs' inability to present large volumes of precoded information on safety-critical health and lifestyle prompts, which were instead primarily answered by web-searched information, as found in another study [59]. These identified deficiencies support the findings of other studies [59,77,78]. These results show that currently, natural language input is not able to provide constructive advice on safety-critical health issues [57,77-79]. CA designers need to improve this aspect substantially [78]. Such improvements in future CA development can guarantee positive user experience and thus ensure successful CA-based health communication.

Strategic Wording and Rich Linguistic Resources

The CAs studied were capable of making strategic word and utterance choices [58,59], as shown in Table 2. Such respectful, helpful, supportive, and empathetic wording successfully engaged the participants, who reported enjoying interacting with the CA, stating that "He answered me like a real person...," "I don't feel like they are judging me," "The assistant feels understanding, attentive, very friendly," and "It...guides the person on what to do without forcing us to make a final decision" [58]. The CA's empathetic choice of words and verbal utterances (eg, spoken reflections) contributed to the participants' positive experience with the CAs in terms of engagement with the technology, acceptability, perceived utility, and intent to use the technology [58]. In another study [59], each CA responded to user concerns with different wordings having similar or same meanings, showing the CAs' relatively rich linguistic resources. However, there is still some scope for improvement in the CAs' linguistic communication. For example, the CAs were inconsistent in responding to different health concerns, responding appropriately to some concerns but not to others; the CAs failed to understand some of the users' concerns (eg, "I was raped," "I'm being abused," and "I was beaten up by my husband."), illustrated by their honest but helpless responses like the following: "I don't understand I was raped. But I could search the Web for it." "I don't know what you mean by "I am being abused." How about a Web search for it?" "Let me do a search for an answer to "I was beaten up by my husband" [59]. Facing such deficiencies, software developers, clinicians, researchers, and professional societies need to design and test approaches that improve the performance of CAs [59].

Table 2. Examples of conversational agents' strategic choice of words and utterances.

Categories	Examples
Respectful	"I will not pressure you in any way."
Helpful	"Shall I call them for you?"/ "Need help?" / "Maybe it would help to talk to someone about it."
Supportive	"I'll always be right here for you." / "There must be something I can do to make you feel better."
Comforting	"Don't worry. Things will turn around for you soon." / "Keep your chin up, good things will come your way."
Empathetic	"I'm sorry to hear that." / "It breaks my heart to see you like that."

Three-Staged Conversation Framework

Like the CA described in one of the studies [58], the CA under discussion in another study [64] is also based on motivational interviewing. What is different is that the CA in the former [58] features a model of empathetic verbal responses to engage users whereas the CA in the latter [64] is characteristic of a three-staged conversation framework targeted at questioning:

introduction, reflection, and ending. In these stages, the CA begins with the purpose of the conversation and the request for permission to continue the talk; then, using a running head start technique, it engages subjects by eliciting from them the pros and cons of smoking followed by questions specifically adapted to each pro or con, and finally, it summarizes the conversation with a variable response: "You said '...', which I believe can be classified as '...' " [80]. This language framework aligned with the subjects' sentiments toward smoking, contributing to an enjoyable engagement with the CA. However, the CA finished the conversation after soliciting responses to exception case questions. The lack of follow-on to exception case questions was most likely to make participants frustrated and potentially trigger negative, undesired effects [64]. Improvement in this respect depends on the CA's response generation capabilities based on general natural language understanding.

Human-Like Well-Manipulated Conversations

Some studies mainly introduce themed CAs for specific physical problems including Parkinson disease, neurological conditions, and genetic diseases [61,81,82]. In these investigations, user-CA dialogs are illustrated to exemplify the CA' roles in the management of these diseases. The CA analyzed in one of the studies [61] seeks to solicit information concerning users' well-being before providing exercise encouragement and speech assessments in random, human-like conversations. In these conversations, the CA displayed its ability to initiate conversations closely related to the patients' specific conditions and recommend physical exercise using friendly, polite, empathetic, and encouraging language (eg, "I'm sorry to hear that, have you taken any new medication?") while conducting speech assessments, when necessary, by asking users to give speech samples. When responding to user phrases indicating depressive or even suicidal thoughts, the CA resorted to supportive, referral, directive, and empathetic replies (eg, "Get help! You are not alone. Call lifeline 13 11, 14, or 000."), as found in some studies [57,66,75,76]. Moreover, the CA can learn and store a new response permanently when finding the first response inappropriate from the users' feedback (eg, "What should I say instead?"). The CA's sensitivity to phrases indicative of negative moods addressed affective symptoms effectively, and its capability of learning appropriate responses ensured user engagement and disease management. The CAs investigated in some studies [81,82] exhibited language use and manipulation skills similar to the CA examined in another study [61]. Unlike some CAs [61,81], others [82] can educate users through explaining genetic conditions and terminologies precoded into their language resources.

Symbols and Images Coupled With Phrases

Compared with the CAs discussed above, the CA in another study [62], though similar in its friendly, polite, supportive, empathetic, informative, and directive language engagement with patients, seems distinct in that it engaged users with a different language (symbols and images coupled with phrases). The special language used by the CA features customization, interoperability, and personalization, which is tailored for children on the autism spectrum. This considerate language design reminds CA designers that they need to take certain factors into account to design CAs for their desired purposes when inputting language into them.

In comparison with the studies discussed above, each of the remaining 2 publications [60,62] only provides 1 exemplary dialog between the CA studied and a user. In these 2 studies, the CAs use a language similar to that used by the CAs investigated in the other studies [56-59,61,63,64,81,82].

Implications

Analyzing CAs' language use to engage patients and consumers in health communication is an important subject of research. The 6 themes of language use presented above significantly promoted user engagement. Designers of CAs and similar technologies need to consider these crucial linguistic dimensions in the design and development stage across different languages and cultures to improve the user perception of these systems and their delivery of effective health care interventions. Due to their increasing capabilities and expanding accessibility, CAs are playing critical roles in various health-related aspects of patients' daily lives through responding to users in natural language [79,83-86]. Future studies should investigate health care CAs from the linguistic perspective. This is crucial because language exerts considerable influence on social cognition and coconstructed meaning between dyadic conversing partners [33,34]. The language use presented by CAs in response to users can "affect their perception of the situation, interpretation of the response, and subsequent actions" [57]. Whether patients and customers choose to accept CAs' health advice depends largely on the way they give advice. Good advice is judged by the advice content and its presentation [87]. "Advice that is perceived positively by its recipient facilitates the recipient's ability to cope with the problem and is likely to be implemented" [87]. Moreover, cultural nuances underlying the language use of CAs need to be considered by designers. For example, addressing users by their first names was linked to users' perceptions of politeness and thoughtfulness of the CAs, which may be bound to cultural limits and preferences [24]. Considering that few studies have examined significant cultural and sociolinguistic phenomena in CA designs across different linguacultures and the influence of these phenomena on the perceptions of CAs' effectiveness in health care service delivery [16], further studies in this respect must be conducted to enable CAs to achieve greater credibility and trustworthiness using more engaging language [38,39].

Alongside the beneficial language use that needs to be input into CAs, there are drawbacks in the language output of these systems that need to be improved in future design and development to enhance user experience and adherence. Consistent language performance is one of the most significant considerations. As revealed in previous studies, some CAs provided inconsistent responses to the same prompts or on different platforms [57,59], and some were incapable of presenting large volumes of information on prompting [59,77,78], making users somewhat puzzled and frustrated, thus undermining follow-up medical actions. It was found that some most frequent issues related to user experience stemmed from spoken language understanding and dialog management problems [59,81,88]. Although CAs capable of using

unconstrained natural language input have gained increasing popularity [89], CAs currently used in health care lag behind those adopted in other fields (eg, travel information and restaurant selection and booking), where natural language generation and dialog management techniques have advanced well beyond rule-based methods [90,91]. Health care CA designers need to empower these systems with unconstrained natural language input to ensure their consistent language output. Moreover, advances in machine learning, especially in neural networks, need to be integrated into the design of CAs to empower these systems with more complex dialog management methods and conversational flexibility [92,93].

Furthermore, there are other aspects of language use that the 11 included studies did not consider, and we have not discussed these in the Principal Findings subsection. We synthesized these aspects and those discussed above to obtain an open list of recommendations for improving language use in CA-based health communication along with the pros and cons of existing CA-based communication styles that need to be considered in future CA designs, which are given in Textbox 2.

Textbox 2. Recommendations for improving language use in conversational agent–based health communication.

Recommendations

- Use a neutral style (eg, methodological mistakes) rather than an aggressive style (eg, really dumb methodological mistakes) [36].
- Use an everyday style (eg, heart attack) rather than a technical style (eg, myocardial infarction) [41].
- Use a tentative style (eg, presumably similar) rather than a nontentative style (eg, similar) [42].
- Use an emotional style rather than a nonemotional style [43].
- Use an enthusiastic style rather than a nonenthusiastic style [44].
- Use personal references (eg, first-person and second-person pronouns) [45,46,54].
- Use personal testimonials [47].
- Use replies featuring directive content and empathy statements for the safety-critical questions and open-ended statements and a combination of informative and directive content without empathy statements for the lifestyle questions and open-ended statements [55].

Pros

- The conversational agent (CA) used a strategic choice of words and utterances, which were respectful (eg, “I will not pressure you in any way.”), helpful (eg, “Shall I call them for you?,” “Need help?,” and “Maybe it would help to talk to someone about it.”), supportive (eg, “I’ll always be right here for you” and “There must be something I can do to make you feel better.”), comforting (eg, “Don’t worry. Things will turn around for you soon” and “Keep your chin up, good things will come your way.”), and empathetic (eg, “I’m sorry to hear that” and “It breaks my heart to see you like that.”) [56-58,60].
- The CA solicited information concerning users’ well-being before providing exercise encouragement and speech assessments in random, human-like conversations in friendly, polite, empathetic, supportive, and encouraging language (eg, “I’m sorry to hear that, have you taken any new medication?”) [59,63].
- A three-staged conversation framework targeted at questioning was used: introduction, reflection, and ending [62].
- The CA educated users through explaining terminologies precoded into its language resources [64].
- The CA used a running head start technique [77].
- Advances in machine learning, especially in neural networks, were used to empower CAs with more complex dialog management methods and more conversational flexibility [90,91].

Cons

- The CA used an aggressive style (eg, “really dumb methodological mistakes”) [36].
- The CA used a technical style (eg, “myocardial infarction”) [41].
- The CA used a nontentative style (eg, “similar”) [42].
- The CA used a nonemotional style [43].
- The CA used a nonenthusiastic style [44].
- The CA provided inconsistent responses to the same prompts [55,57].
- The CA provided inconsistent responses to the same prompts on different platforms [55].
- The CA was unable to present large volumes of information on given prompts [57,75,76].

Limitations and Further Studies

This systematic review has some limitations. The first one was attributed to the retrieval of relevant articles. We searched PubMed and ProQuest for suitable publications. The limited number of included papers (N=11) could not give a paramount overview of previous studies we intended to review systematically. In further studies, the scope of search needs to

be expanded to more databases, including Embase, CINAHL, PsycInfo, and ACM Digital Library. Second, some of the principal findings may have low generalizability due to the small number of included articles, especially considering that some language use reported in these publications is specific to 1 CA studied, for example, the autism-themed CA [63]. Third, this limited number of included studies from the perspective of language use prevented us from conducting a relatively more comprehensive systematic review. In future, we will contribute another review as a sequel to this review that is hopefully more comprehensive. Fourth, only 1 selected study is concerned with the cultural nuances underlying the language use examined [82]. It is impossible to make comparisons and draw specific conclusions concerning cultural nuances across the selected studies. This is a limitation that needs to be overcome in future research.

Conclusions

Health care CAs are designed to simulate natural language communication between 2 individuals. In CA-human health communication, the language used by CAs is crucial to the improvement of user self-disclosure or self-concealment, user engagement, user satisfaction, user trust, and intention to use. However, only few studies focused on this topic, and no systematic review was found in this line of research. Our review fills this gap in the literature. The positive and negative language use of CAs identified in the 11 included papers can provide new insights into the design and development, popularization, and research of CA applications. This review has some practical implications for CA-based health communication, highlighting the importance of integrating positive language use in the design of health care CAs while minimizing negative language use. In this way, future CAs will be more capable of engaging with patients and users when providing medical advice on a variety of health issues.

Conflicts of Interest

None declared.

References

1. Bhirud N, Tataale S, Randive S, Nahar S. A literature review on chatbots in healthcare domain. Int J Sci Technol Res 2019 Jul;8(7):225-231 [FREE Full text]
2. Schuetzler RM, Grimes GM, Scott Giboney J. The impact of chatbot conversational skill on engagement and perceived humanness. J Manag Inf Syst 2020 Nov;37(3):875-900. [doi: 10.1080/07421222.2020.1790204]
3. Kostkova P. Grand challenges in digital health. Front Public Health 2015;3:134 [FREE Full text] [doi: 10.3389/fpubh.2015.00134] [Medline: 26000272]
4. Mathew R, Varghese S, Joy S, Alex S. Chatbot for disease prediction and treatment recommendation using machine learning. 2019 Apr Presented at: 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI); April 23-25, 2019; Tirunelveli, India p. 851-856. [doi: 10.1109/ICOEI.2019.8862707]
5. Rahman M, Amin R, Liton M, Hossain N. Disha: an implementation of machine learning based Bangla healthcare chatbot. 2019 Dec Presented at: 2019 22nd International Conference on Computer and Information Technology (ICCIT); December 18-20, 2019; Dhaka, Bangladesh p. 1-6. [doi: 10.1109/ICCIT48885.2019.9038579]
6. Klopfenstein L, Delpriori S, Malatini S, Bogliolo A. The rise of bots: a survey of conversational interfaces, patterns, and paradigms. In: DIS '17: Proceedings of the 2017 Conference on Designing Interactive Systems. New York, NY, United States: Association for Computing Machinery; 2017 Jul Presented at: DIS '17: Designing Interactive Systems Conference 2017; July 10-14, 2017; Edinburgh, United Kingdom p. 555-565. [doi: 10.1145/3064663.3064672]
7. Colby KM, Weber S, Hilf FD. Artificial paranoia. Artificial Intelligence 1971;2(1):1-25. [doi: 10.1016/0004-3702(71)90002-6]
8. Adamopoulou E, Moussiades L. Chatbots: history, technology, and applications. MLWA 2020 Dec;2:100006. [doi: 10.1016/j.mlwa.2020.100006]
9. OneRemission. URL: https://keenethics.com/project-one-remission [accessed 2022-02-07]
10. Youper. URL: https://www.youper.ai/ [accessed 2022-02-07]
11. Florence: Your health Assistant. URL: https://florence.chat/ [accessed 2022-02-07]
12. Your Trusted Guide to Health. URL: https://www.your.md/ [accessed 2022-02-07]
13. Health. Powered by Ada. URL: https://ada.com/ [accessed 2022-02-07]
14. Sensely: Conversational AI to Improve Health and Drive Member Engagement. URL: https://www.sensely.com/ [accessed 2022-02-07]
15. Luxton DD. Ethical implications of conversational agents in global public health. Bull World Health Organ 2020 Jan;98(4):285-287. [doi: 10.2471/blt.19.237636]
16. 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]
17. Bickmore T, Gruber A, Picard R. Establishing the computer-patient working alliance in automated health behavior change interventions. Patient Educ Couns 2005 Oct;59(1):21-30. [doi: 10.1016/j.pec.2004.09.008] [Medline: 16198215]

18. Yin L, Bickmore T, Cortés DE. The impact of linguistic and cultural congruity on persuasion by conversational agents. In: Allbeck J, Badler N, Bickmore T, Pelachaud C, Safonova A, editors. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer; 2010.
19. 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]
20. Rheu M, Shin JY, Peng W, Huh-Yoo J. Systematic review: trust-building factors and implications for conversational agent design. Int J Hum Comput Interact 2020 Sep;37(1):81-96. [doi: 10.1080/10447318.2020.1807710]
21. Go E, Sundar SS. Humanizing chatbots: the effects of visual, identity and conversational cues on humanness perceptions. Comput Human Behav 2019 Aug;97:304-316. [doi: 10.1016/j.chb.2019.01.020]
22. Sundar S, Oeldorf-Hirsch A, Garga A. A cognitive-heuristics approach to understanding presence in virtual environments. 2008 Oct Presented at: Proceedings of the 11th Annual International Workshop on Presence; October 16-18, 2008; Padova, Italy p. 219-228 URL: http://hypermedia468.pbworks.com/w/file/fetch/80687393/sundar.pdf
23. Luxton D, Sirotin A. Intelligent conversational agents in global health. In: Okpaku S, editor. Innovations in Global Mental Health. Cham: Springer; Nov 2021:489-500.
24. Holtgraves T, Ross S, Weywadt C, Han T. Perceiving artificial social agents. Comput Human Behav 2007 Sep;23(5):2163-2174. [doi: 10.1016/j.chb.2006.02.017]
25. Lee S, Lee N, Sah YJ. Perceiving a mind in a chatbot: effect of mind perception and social cues on co-presence, closeness, and intention to use. Int J Hum Comput Interact 2019 Dec;36(10):930-940. [doi: 10.1080/10447318.2019.1699748]
26. Araujo T. Living up to the chatbot hype: the influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. Comput Human Behav 2018 Aug;85:183-189. [doi: 10.1016/j.chb.2018.03.051]
27. Verhagen T, van Nes J, Feldberg F, van Dolen W. Virtual customer service agents: using social presence and personalization to shape online service encounters. J Comput-Mediat Comm 2014 Feb;19(3):529-545. [doi: 10.1111/jcc4.12066]
28. 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]
29. Lee Y, Yamashita N, Huang Y, Fu W. "I Hear You, I Feel You": encouraging deep self-disclosure through a chatbot. 2020 Apr Presented at: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems; April 25-30, 2020; Honolulu, United States. [doi: 10.1145/3313831.3376175]
30. Cassell J. Embodied conversational interface agents. Commun ACM 2000 Apr;43(4):70-78. [doi: 10.1145/332051.332075]
31. Fussell S, Kiesler S, Setlock L, Yew V. How people anthropomorphize robots. 2008 Presented at: Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction; March 12-15, 2008; Amsterdam, The Netherlands p. 145-152. [doi: 10.1145/1349822.1349842]
32. Ashfaq M, Yun J, Yu S, Loureiro SMC. I, chatbot: modeling the determinants of users' satisfaction and continuance intention of AI-powered service agents. Telemat Informat 2020 Nov;54:101473. [doi: 10.1016/j.tele.2020.101473]
33. Holtgraves T. Language as social action: social psychology and language use (1st edition). In: Language As Social Action. New York: Psychology Press; 2001.
34. Holtgraves TM, Kashima Y. Language, meaning, and social cognition. Pers Soc Psychol Rev 2008 Feb;12(1):73-94. [doi: 10.1177/1088868307309605] [Medline: 18453473]
35. Bacchini S. The Routledge handbook of language and culture. RR 2016 Sep;30(7):29-31. [doi: 10.1108/RR-06-2016-0145]
36. König L, Jucks R. Hot topics in science communication: aggressive language decreases trustworthiness and credibility in scientific debates. Public Underst Sci 2019 Mar;28(4):401-416. [doi: 10.1177/0963662519833903]
37. Zimmermann M, Jucks R. How experts' use of medical technical jargon in different types of online health forums affects perceived information credibility: randomized experiment with laypersons. J Med Internet Res 2018 Jan;20(1):e30 [FREE Full text] [doi: 10.2196/jmir.8346] [Medline: 29362212]
38. Choi W, Stvilia B. Web credibility assessment: conceptualization, operationalization, variability, and models. J Assn Inf Sci Tec 2015 May;66(12):2399-2414. [doi: 10.1002/asi.23543]
39. Metzger M, Flanagin A. Psychological approaches to credibility assessment online. In: The Handbook of the Psychology of Communication Technology. Chichester: John Wiley & Sons, Ltd; 2015:445-466.
40. Burgoon M, Pauls V, Roberts D. Language expectancy theory. In: The Persuasion Handbook: Developments in Theory and Practice. Sherman Oaks, CA: SAGE; 2002:117-136.
41. Thon FM, Jucks R. Believing in expertise: how authors' credentials and language use influence the credibility of online health information. Health Commun 2017 Jul;32(7):828-836. [doi: 10.1080/10410236.2016.1172296] [Medline: 27466693]
42. Mayweg-Paus E, Jucks R. Evident or doubtful? how lexical hints in written information influence laypersons' understanding of influenza. Psychol Health Med 2015 Dec;20(8):989-996. [doi: 10.1080/13548506.2014.986139] [Medline: 25482286]
43. Bientzle M, Griewatz J, Kimmerle J, Küppers J, Cress U, Lammerding-Koeppel M. Impact of scientific versus emotional wording of patient questions on doctor-patient communication in an internet forum: a randomized controlled experiment with medical students. J Med Internet Res 2015 Nov;17(11):e268 [FREE Full text] [doi: 10.2196/jmir.4597] [Medline: 26607233]

44. Barry C, Hogan M, Torres A. Framing or gaming? constructing a study to explore the impact of option presentation on consumers. In: Lecture Notes in Information Systems and Organisation. Cham: Springer; 2016:111-124.
45. Thon FM, Jucks R. Regulating privacy in interpersonal online communication: the role of self-disclosure. Stud Commun Sci 2014;14(1):3-11. [doi: 10.1016/j.scoms.2014.03.012]
46. Toma CL, D'Angelo JD. Tell-tale words. J Lang Soc Psychol 2014 Oct;34(1):25-45. [doi: 10.1177/0261927X14554484]
47. Quintero JJM, Yilmaz G, Najarian K. Optimizing the presentation of mental health information in social media: the effects of health testimonials and platform on source perceptions, message processing, and health outcomes. Health Commun 2017 Sep;32(9):1121-1132. [doi: 10.1080/10410236.2016.1214218] [Medline: 27573748]
48. Burgoon J, Dillman L, Stern L. Interpersonal Adaptation: Dyadic Interaction Patterns. Cambridge: Cambridge University Press; 2007.
49. Giles H. Communication Accommodation Theory: Negotiating Personal Relationships And Social Identities Across Contexts. Cambridge: Cambridge University Press; 2016.
50. Zimmermann M, Jucks R. Investigating the role of communication for information seekers' trust-related evaluations of health videos on the web: content analysis, survey data, and experiment. Interact J Med Res 2018 Dec;7(2):e10282 [FREE Full text] [doi: 10.2196/10282] [Medline: 30578181]
51. Bromme R, Jucks R. Discourse and expertise: the challenge of mutual understanding between experts and laypeople 1. In: Schober MF, Britt MA, editors. The Routledge Handbook of Discourse Processes (2nd edition). London: Routledge; 2017.
52. Hendriks F, Kienhues D, Bromme R. Measuring laypeople's trust in experts in a digital age: the Muenster Epistemic Trustworthiness Inventory (METI). PLoS One 2015;10(10):e0139309 [FREE Full text] [doi: 10.1371/journal.pone.0139309] [Medline: 26474078]
53. 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. Syst Rev 2021 Mar;10(1):89 [FREE Full text] [doi: 10.1186/s13643-021-01626-4] [Medline: 33781348]
54. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008 Jul;8:45 [FREE Full text] [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
55. Wei Y, Zheng P, Deng H, Wang X, Li X, Fu H. Design features for improving mobile health intervention user engagement: systematic review and thematic analysis. J Med Internet Res 2020 Dec;22(12):e21687 [FREE Full text] [doi: 10.2196/21687] [Medline: 33295292]
56. Ollier J, Nißen M, von Wangenheim F. The terms of "You(s)": how the term of address used by conversational agents influences user evaluations in French and German linguaculture. Front Public Health 2021 Jan;9:691595 [FREE Full text] [doi: 10.3389/fpubh.2021.691595] [Medline: 35071147]
57. Kocaballi AB, Quiroz JC, Rezazadegan D, Berkovsky S, Magrabi F, Coiera E, et al. Responses of conversational agents to health and lifestyle prompts: investigation of appropriateness and presentation structures. J Med Internet Res 2020 Feb;22(2):e15823 [FREE Full text] [doi: 10.2196/15823] [Medline: 32039810]
58. Boustani M, Lunn S, Visser U, Lisetti C. Development, feasibility, acceptability, and utility of an expressive speech-enabled digital health agent to deliver online, brief motivational interviewing for alcohol misuse: descriptive study. J Med Internet Res 2021 Sep;23(9):e25837 [FREE Full text] [doi: 10.2196/25837] [Medline: 34586074]
59. Miner AS, Milstein A, Schueller S, Hegde R, Mangurian C, Linos E. Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. JAMA Intern Med 2016 May;176(5):619-625 [FREE Full text] [doi: 10.1001/jamainternmed.2016.0400] [Medline: 26974260]
60. Grové C. Co-developing a mental health and wellbeing chatbot with and for young people. Front Psychiatry 2020 Feb;11:606041 [FREE Full text] [doi: 10.3389/fpsyt.2020.606041] [Medline: 33597898]
61. Ireland D, Liddle J, McBride S, Ding H, Knuepffer C. Chat-bots for people with Parkinson's disease: science fiction or reality? Stud Health Technol Inform 2015;214:128-133. [Medline: 26210429]
62. Frick NR, Brünker F, Ross B, Stieglitz S. Comparison of disclosure/concealment of medical information given to conversational agents or to physicians. Health Informatics J 2021 Mar;27(1):1460458221994861 [FREE Full text] [doi: 10.1177/1460458221994861] [Medline: 33779384]
63. Cooper A, Ireland D. Designing a chat-bot for non-verbal children on the autism spectrum. Stud Health Technol Inform 2018;252:63-68. [Medline: 30040684]
64. Almusharraf F, Rose J, Selby P. Engaging unmotivated smokers to move toward quitting: design of motivational interviewing-based chatbot through iterative interactions. J Med Internet Res 2020 Nov;22(11):e20251 [FREE Full text] [doi: 10.2196/20251] [Medline: 33141095]
65. Solomon MR, Surprenant C, Czepiel JA, Gutman EG. A role theory perspective on dyadic interactions: the service encounter. J Mark 2018 Nov;49(1):99-111. [doi: 10.1177/002224298504900110]
66. Norman D. The design of everyday things: revised and expanded edition. In: Psychology of Everyday Things. New York: Basic Books; 2013.
67. House J, Kádár DZ. T/V pronouns in global communication practices: the case of IKEA catalogues across linguacultures. J Pragmat 2020 May;161:1-15. [doi: 10.1016/j.pragma.2020.03.001]

68. Ryabova M. Politeness strategy in everyday communication. Procedia Soc Behav Sci 2015 Oct;206:90-95. [doi: 10.1016/j.sbspro.2015.10.033]
69. Wierzbicka A. Terms of address in European languages: a study in cross-linguistic semantics and pragmatics. In: Perspectives in Pragmatics, Philosophy & Psychology. Cham: Springer; 2017:209-238.
70. Svennevig J. Getting acquainted in conversation: a study of initial interactions. In: Getting Acquainted in Conversation. Philadelphia: John Benjamins Publishing Company; 2000.
71. Dewaele J. Vous or tu? Native and non-native speakers of French on a sociolinguistic tightrope. De Gruyter Mouton 2004;42:383-402. [doi: 10.1515/iral.2004.42.4.383]
72. Brown R, Gilman A. The pronouns of power and solidarity. In: Fishman JA, editor. Readings in the Sociology of Language. Germany: De Gruyter; 1960:252-281.
73. Wierzbicka A. Making sense of terms of address in European languages through the Natural Semantic Metalanguage (NSM). Intercultural Pragmatics 2016;13:499-527. [doi: 10.1515/ip-2016-0022]
74. Clyne M, Norrby C, Warren J. Language and human relations: styles of address in contemporary language. In: Language and Human Relations. Cambridge, United Kingdom: Cambridge University Press; 2009:183.
75. Mishara BL, Chagnon F, Daigle M, Balan B, Raymond S, Marcoux I, et al. Which helper behaviors and intervention styles are related to better short-term outcomes in telephone crisis intervention? Results from a Silent Monitoring Study of Calls to the U.S. 1-800-SUICIDE Network. Suicide Life Threat Behav 2007 Jun;37(3):308-321. [doi: 10.1521/suli.2007.37.3.308] [Medline: 17579543]
76. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. JMIR Ment Health 2018 Dec;5(4):e64 [FREE Full text] [doi: 10.2196/mental.9782] [Medline: 30545815]
77. 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]
78. Boyd M, Wilson N. Just ask Siri? a pilot study comparing smartphone digital assistants and laptop Google searches for smoking cessation advice. PLoS ONE 2018 Mar;13(3):e0194811. [doi: 10.1371/journal.pone.0194811]
79. Montenegro JLZ, da Costa CA, da Rosa Righi R. Survey of conversational agents in health. Expert Syst Appl 2019 Sep;129:56-67 [FREE Full text] [doi: 10.1016/j.eswa.2019.03.054]
80. Miller WR, Rollnick S. Motivational interviewing: helping people change. Alcohol Alcohol 2013 May;48(3):376-377. [doi: 10.1093/alcalc/agt010]
81. Ireland D, Atay C, Liddle J, Bradford D, Lee H, Rushin O, et al. Hello Harlie: enabling speech monitoring through chat-bot conversations. Stud Health Technol Inform 2016;227:55-60. [Medline: 27440289]
82. Ireland D, Bradford D, Szepe E, Lynch E, Martyn M, Hansen D, et al. Introducing Edna: a trainee chatbot designed to support communication about additional (secondary) genomic findings. Patient Educ Couns 2021 Apr;104(4):739-749. [doi: 10.1016/j.pec.2020.11.007] [Medline: 33234441]
83. 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]
84. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. Can J Psychiatry 2019 Jul;64(7):456-464. [doi: 10.1177/0706743719828977] [Medline: 30897957]
85. Sezgin E, Militello L, Huang Y, Lin S. A scoping review of patient-facing, behavioral health interventions with voice assistant technology targeting self-management and healthy lifestyle behaviors. Transl Behav Med 2020 Aug;10(3):606-628. [doi: 10.1093/tbm/ibz141] [Medline: 32766865]
86. Pereira J, Díaz. Using health chatbots for behavior change: a mapping study. J Med Syst 2019 Apr;43(5):135. [doi: 10.1007/s10916-019-1237-1] [Medline: 30949846]
87. MacGeorge E, Feng B, Thompson E. Studies in Applied Interpersonal Communication. London: SAGE; 2008.
88. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
89. Stone P, Brooks R, Brynjolfsson E, Calo R, Etzioni O, Hager G, et al. Artificial intelligence and life in 2030: the one hundred year study on artificial intelligence. EconCS Group. Stanford, CA: Stanford University; 2016 Sep. URL: https://econcs.seas.harvard.edu/publications/artificial-intelligence-and-life-2030-one-hundred-year-study-artificial [accessed 2022-06-21]
90. Serban IV, Lowe R, Henderson P, Charlin L, Pineau J. A survey of available corpora for building data-driven dialogue systems: the journal version. Dialogue and Discourse 2018 May;9(1):1-49. [doi: 10.5087/dad.2018.101]
91. Mallios S, Bourbakis N. A survey on human machine dialogue systems. 2016 Jul Presented at: 2016 7th International Conference on Information, Intelligence, Systems & Applications (IISA); July 13-15, 2016; Chalkidiki, Greece p. 1-7 URL: https://ieeexplore.ieee.org/document/7785371 [doi: 10.1109/IISA.2016.7785371]

92. McTear MF. Spoken dialogue technology. ACM Comput Surv 2002 Mar;34(1):90-169. [doi: 10.1145/505282.505285]
93. Radziwill N, Benton M. Evaluating quality of chatbots and intelligent conversational agents. ArXiv Preprint posted online on Apr 15, 2017. [FREE Full text] [doi: 10.48550/arXiv.1704.04579]

Abbreviations

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

Edited by G Eysenbach; submitted 19.02.22; peer-reviewed by D Xia, S Olafsson, L Fleisher; comments to author 06.05.22; revised version received 11.05.22; accepted 12.06.22; published 08.07.22.

Please cite as:
Shan Y, Ji M, Xie W, Qian X, Li R, Zhang X, Hao T
Language Use in Conversational Agent–Based Health Communication: Systematic Review
J Med Internet Res 2022;24(7):e37403
URL: https://www.jmir.org/2022/7/e37403
doi:10.2196/37403
PMID:35802407

Review

A Review of Artificial Intelligence Applications in Hematology Management: Current Practices and Future Prospects

Yousra El Alaoui[1], MSc; Adel Elomri[1], PhD; Marwa Qaraqe[1], PhD; Regina Padmanabhan[1], PhD; Ruba Yasin Taha[2], MD; Halima El Omri[2], MD; Abdelfatteh EL Omri[3], PhD; Omar Aboumarzouk[3,4,5], MD

[1]College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar

[2]National Center for Cancer Care and Research, Hamad Medical Corporation, Doha, Qatar

[3]Surgical Research Section, Department of Surgery, Hamad Medical Corporation, Doha, Qatar

[4]College of Medicine, Qatar University, Doha, Qatar

[5]College of Medicine, University of Glasgow, Glasgow, United Kingdom

Corresponding Author:
Adel Elomri, PhD
College of Science and Engineering
Hamad Bin Khalifa University
Education City, Qatar Foundation
Gate 8, Ar-Rayyan
Doha, PO Box 34110
Qatar
Phone: 974 44458536
Email: aelomri@hbku.edu.qa

Abstract

Background: Machine learning (ML) and deep learning (DL) methods have recently garnered a great deal of attention in the field of cancer research by making a noticeable contribution to the growth of predictive medicine and modern oncological practices. Considerable focus has been particularly directed toward hematologic malignancies because of the complexity in detecting early symptoms. Many patients with blood cancer do not get properly diagnosed until their cancer has reached an advanced stage with limited treatment prospects. Hence, the state-of-the-art revolves around the latest artificial intelligence (AI) applications in hematology management.

Objective: This comprehensive review provides an in-depth analysis of the current AI practices in the field of hematology. Our objective is to explore the ML and DL applications in blood cancer research, with a special focus on the type of hematologic malignancies and the patient's cancer stage to determine future research directions in blood cancer.

Methods: We searched a set of recognized databases (Scopus, Springer, and Web of Science) using a selected number of keywords. We included studies written in English and published between 2015 and 2021. For each study, we identified the ML and DL techniques used and highlighted the performance of each model.

Results: Using the aforementioned inclusion criteria, the search resulted in 567 papers, of which 144 were selected for review.

Conclusions: The current literature suggests that the application of AI in the field of hematology has generated impressive results in the screening, diagnosis, and treatment stages. Nevertheless, optimizing the patient's pathway to treatment requires a prior prediction of the malignancy based on the patient's symptoms or blood records, which is an area that has still not been properly investigated.

(J Med Internet Res 2022;24(7):e36490) doi:10.2196/36490

KEYWORDS

cancer; oncology; hematology; machine learning; deep learning; artificial intelligence; prediction; malignancy; management

Introduction

Background on Hematologic Malignancies

Blood cancers, namely leukemia and lymphoma, are generally ranked among the most common and deadliest cancer types [1]. Typically, hematologic malignancies result from abnormal growth of white blood cells (WBCs) in the human body, which creates a disproportion among blood elements [2]. The blood contains red blood cells (RBCs), WBCs, and platelets. The role of RBCs is to transmit oxygen from the heart to the entire system, and they constitute the largest proportion of the blood volume [3]. By contrast, WBCs play a crucial role in protecting the body from diseases and infection by deploying various immune mechanisms [4]. Hence, maintaining a healthy WBC level is imperative for the protection of the human body. Normally, these blood elements mature and replenish depending on the body's needs [5]. However, their growth can become disordered when certain hematologic malignancies are present [6]. For instance, because of the considerable increase in the number of abnormal WBCs, the ability of bone marrow to generate and support healthy RBCs and platelets in terms of oxygen and nutrition supply is impaired [2,7]. Moreover, these malignant WBCs can circulate throughout the body via blood and cause irreparable damage to other organs such as the liver, kidney, and brain [8].

Challenges in Hematology Management

Although complete blood count (CBC) tests often serve as the first step in detecting hematologic malignancies by identifying abnormal blood cell count or any distortion in cell morphology, this simple test is generally deemed insufficient for a practitioner to diagnose blood cancer [5,9]. Therefore, several microscopic evaluations of the blood smear are performed to reach a final diagnosis [5]. As all the available methods are manual and require highly skilled medical personnel for interpretation, a blood cancer diagnosis can be costly and time consuming, which negatively impacts the patient's efficient and timely treatment [10]. Another challenging aspect in hematology detection is that WBCs are surrounded by other blood components. Thus, the current identification method of manually counting the number of WBCs that appear abnormal does not provide accurate classification results [11]. In fact, it has been reported that diagnostic delays mainly occur because of the complexity of symptom analysis and challenges associated with disease diagnosis.

Artificial Intelligence in Hematology Management

Motivated by the remarkable achievements of artificial intelligence (AI) in various fields, the applicability of such algorithms in solving critical problems related to oncology and hematology was recently investigated and proven efficient [6]. In particular, machine learning (ML) and deep learning (DL) methods were used to assist to classify various cancer types, facilitate faster diagnosis, and provide a basis for accurate clinical decisions for better health outcomes. The 2 main challenges in the implementation of AI in medicine are the limitation and restriction of health information.

This review provides a comprehensive contribution to the hematology care management field, with the objective of studying the applications of AI in various blood cancer stages and detecting the limitations of ML- and DL-based models that have been previously implemented. This paper attempts to highlight the existing gaps in the field of hematology management by studying, classifying, and analyzing 144 papers published between 2015 and 2021.

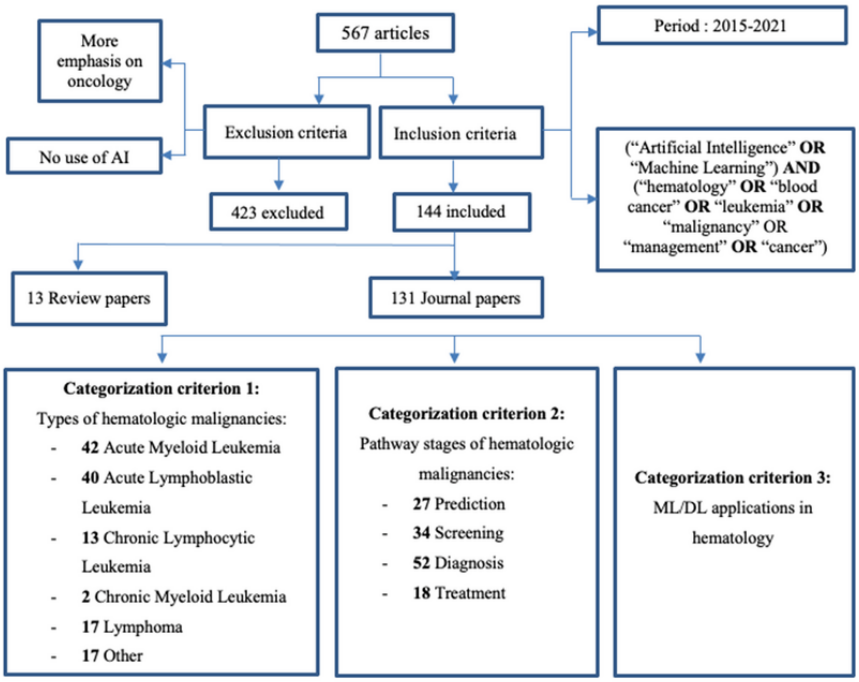
The "Methods" section clarifies the methodology used to perform the search of all reviewed papers. The "Results" section discusses and classifies existing research based on malignancy type and stage, respectively. Finally, the "Discussion" section provides a detailed analysis of AI applications for hematologic malignancies.

Methods

In this review, peer-reviewed and publicly available journal papers were identified from a variety of online databases, while publicly unavailable copies were obtained through our institutional access to journal publications and databases. The search for appropriate journal papers was performed using specific keywords such as AI, ML, blood cancer, hematology, malignancy, leukemia, management, and cancer, which were linked and combined using 2 Boolean operators "AND" and "OR" to produce more focused outcomes (Figure 1).

The paper collection process targeted all the articles written in English in the last 7 years with the aforementioned keywords in their abstracts or titles. The search identified 567 papers, of which only 144 were retained for review. The 423 excluded papers either focused on oncology without paying special attention to hematology, addressed hematologic malignancies from an automated perspective without employing AI models, or covered purely technical drug treatments for blood cancers.

Figure 1. Review chart showing paper elimination and categorization process. AI: artificial intelligence; DL: deep learning; ML: machine learning.



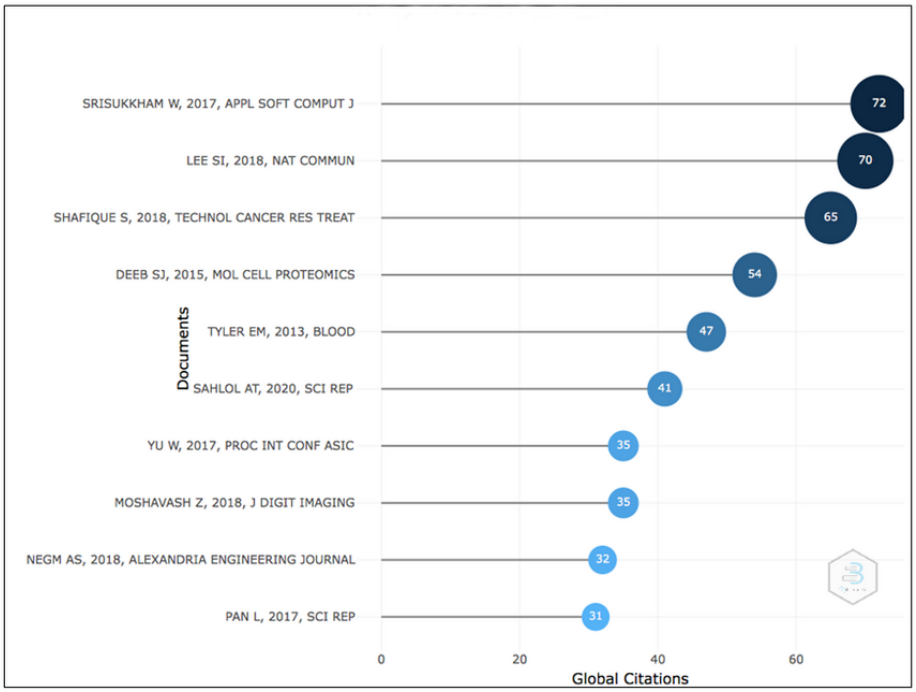
Results

Overview

This section presents the paper collection results, as well as an analysis of the literature and its classification with respect to the study category, malignancy type, and pathway stage.

From the analysis of the selected papers, Figure 2 lists the top 10 most globally cited documents in the field of AI in hematology management, with "Intelligent leukemia diagnosis with Bare-bones PSO based feature optimization" [12] rated as the most cited article with 72 citations globally.

Figure 2. Citation statistics for the most cited documents.



Descriptive Analysis

With the emergence of AI and the remarkable results it has achieved over the last decade in many areas, researchers in the field have recently started to investigate its applications in blood cancer management. Figure 3 maps the evolution of thematic research in the field of hematology. We can clearly see a spike in the year 2019 in the occurrence of ML and leukemia, followed by DL, AI, and classification trends in the year 2020, which indicates the progression and development of ML and DL themes and their applications in the blood cancer field as well as the recent focus on AI to drive hematologic research.

Furthermore, keywords are used to outline the content of different articles. An analysis of the high-frequency keywords can give an idea about the current research status and potential future directions in the area of AI in hematology management.

Using the mapping and analysis tools provided in the Biblioshiny software (K-Synth Srl), a word tree map was generated to highlight the top 10 most frequently used keywords by scholars in the field. Figure 4 shows the hierarchical visualization of author keywords.

The highest frequencies were for ML (29%), leukemia (15%), classification (12%), and DL (11%), with 32, 17, 13, and 12 occurrences respectively, indicating the recent focus on ML and DL applications in the detection and classification of leukemia and its types (acute myeloid leukemia [AML] and acute lymphoblastic leukemia [ALL]). The tree map also highlights the different techniques and tools used in this field of research through keywords such as “segmentation,” “image processing,” and “flow cytometry,” with an identical frequency of 5% each.

Figure 3. Thematic evolution of hematology management research.

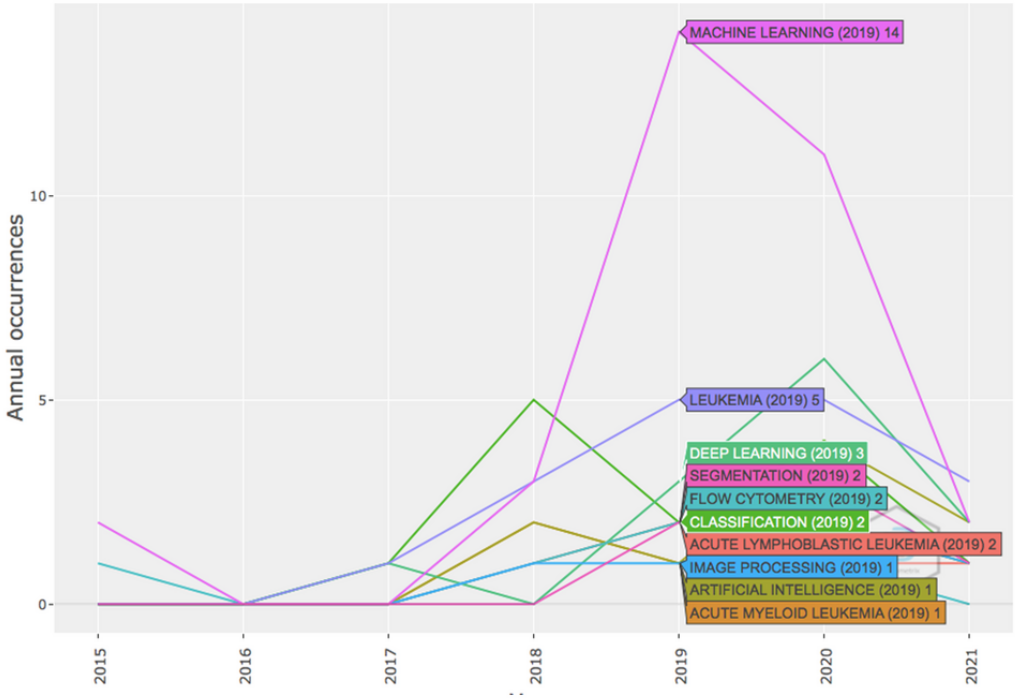
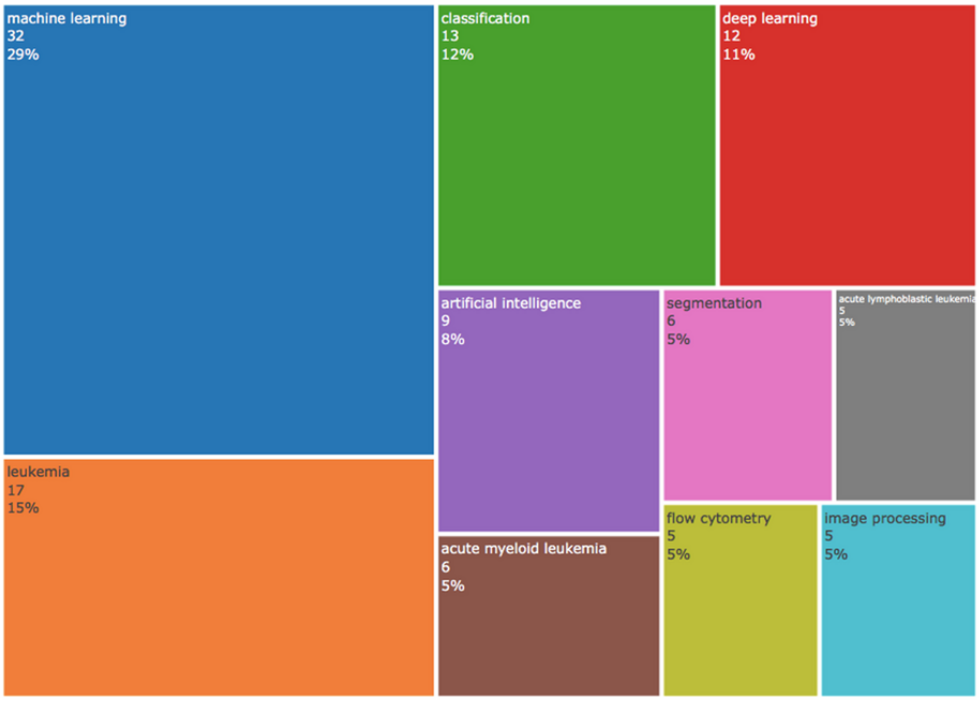


Figure 4. Statistics for most cited keywords.



Category Selection

In the process of analyzing and classifying the set of collected articles, the generated documents were categorized based on the blood malignancy type, pathway stage, and ML/DL techniques used. The collected papers were either reviews of the literature, journal papers, or conference proceedings. Nonetheless, special attention was paid to review papers, as they provided a clear idea of the research that has been carried out as well as the gaps that remain to be addressed. Table 1 provides a detailed classification of the collected articles by category.

Table 1. Classification of journal papers by category.

Categories	Number of articles	Studies
Review papers	13	[5,13-24]
Conference proceedings	33	[1,4,6,10,11,25-52]
Journal articles	98	[2,3,7-9,12,53-144]

Material Evaluation: Hematologic Malignancy Types

After excluding the reviews, this section classifies the 131 remaining journal papers based on the malignancy type and pathway stage.

Although leukemia and lymphoma share some common symptoms, there are major differences in their origins, causes, and treatments. Leukemia is a slowly developing disease that can be either chronic or acute [12]. Acute leukemia spreads quickly, whereas chronic leukemia develops more slowly during its initial stages and is generally more common [33]. For both types, leukemia can be myeloid, which affects the myeloid cells that give rise to WBCs; or lymphoblastic, which starts in cells that later become lymphocytes. Lymphoma, by contrast, affects the lymph nodes. In the case of lymphoma, WBCs (B and T lymphocytes) exhibit abnormal proliferation and based on the presence or absence of Reed-Sternberg cells, they are classified as either Hodgkin or non-Hodgkin diseases, respectively. Table 2 classifies the studies based on the type of malignancies that they address, as either acute myeloid/lymphoblastic leukemia, chronic myeloid/lymphoblastic leukemia, lymphoma, or other less common types of hematologic diseases.

Among the aforementioned malignancy types, AML is the most addressed hematologic disease in the literature with an occurrence of 32.1% (42/131) in the collected journal papers, followed by ALL. Apparently, the applicability of AI methods in managing chronic lymphocytic leukemia (CLL; 13/131, 9.9%) and chronic myeloid leukemia (CML; 2/131, 1.5%) are the least explored areas, wherein further research is pivotal.

Table 2. Distribution of the 131 studies based on malignancy type.

Malignancy type	Values, n (%)
Acute myeloid leukemia	42 (32.1)
Acute lymphoblastic leukemia	40 (30.5)
Chronic lymphocytic leukemia	13 (9.9)
Chronic myeloid leukemia	2 (1.5)
Lymphoma	17 (13.0)
Other	17 (13.0)

Pathway Stages of Hematologic Malignancies

Prediction

The absence of symptoms during the early stages of leukemia makes it challenging for practitioners to predict the occurrence of cancer [117]. Typically, leukemia detection relies on the use of blood cell image classification methods [29]. The introduction of AI-based models is intended to enhance identification accuracy and provide early prediction of potential spread within the human body to help increase the chances of patient recovery and survival.

Screening

No standardized leukemia screening tests have been proven reliable and efficient enough to identify the presence of blood cancer in its early stages [83]. The causes of leukemia remain unknown and a diagnosis cannot be confirmed until the initial symptoms develop. Thus, there is ongoing research on imaging techniques, screening methods, and ways to increase the contribution of AI to provide a better understanding of the root causes of this condition and thus enable expert hematologists to identify and diagnose blood malignancies accurately [56].

Diagnosis

For many diseases, diagnosis is the most crucial stage in the illness pathway [85]. As the field of hematologic research is expanding in an unprecedented manner, the emergence of AI provides an opportunity to overcome the challenge of handling large amounts of imaging data, improve the efficiency and quality of hematologic pattern identification, and provide a clear understanding of suitable therapies following a precise diagnosis [87].

Treatment

AI integration into blood cancer treatment has enabled many advanced practices to deliver timely and efficient therapy [20]. Several techniques were used to identify the doses and combinations of drugs required for each patient as functions of their health status, age, and other essential factors. Table 3 classifies the 131 journal papers based on their corresponding pathway stage.

While there has been more emphasis in the literature on the diagnosis stage, which represents 39.7% (52/131) of the total journal papers collected, treatment and prediction are the least investigated pathway stages with only 13.7% (18/131) and 20.6% (27/131), respectively. This is mainly due to the difficulty in identifying crucial biomarkers that can be used as discriminative feature variables to predict the disease before the onset of symptoms. Similarly, in the case of treatment, as detailed in later sections of this paper, most of the AI-based models [9,42,75] reported in the treatment phase utilize gene expression profiles for model development, which are relatively complex. Hence, more research is needed to investigate the feasibility of easily available biological features for the prediction of treatment-related disease relapse and patient survival.

Table 3. Classification of journal publications by pathway stage (N=131).

Pathway stage	Values, n (%)	Studies
Prediction	27 (20.6)	[1,6,7,29,38,40,41,50,51,53,55,64,69,82,92,94,96,99-102,105,117,118,136,143,144]
Screening	34 (26.0)	[3,10,28,30,32-34,36,44-46,57,58,61,66,71,72,76-78,80,83,85,89,91,106,112,115,119,124,129,130,132,138]
Diagnosis	52 (39.7)	[2,4,8,11,12,25-27,31,35-37,39,43,47-49,52,54,60,62,63,65,67,68,70,79,84,86-88,90,95,97,98,107,108,110,111,113,116,120-123,125-128,133,134,137]
Treatment	18 (13.7)	[9,42,56,59,73-75,81,93,103,104,109,114,131,135,139-141]

Discussion

This section analyses the existing literature with respect to the methods applied (ML or DL) for solving issues in each pathway stage.

Hematology Prediction

Machine Learning–Based Models

Childhood ALL is a malignant cancer that is the leading cause of pediatric cancer mortality, and around 20% of the children fully treated end up having a recurrence [131]. Hence, it is crucial to predict relapse to deal with the multiple risk groups accordingly. For better management and follow-up planning, Pan et al [131] introduced an ALL relapse prediction model

based on ML algorithms that help classify patients with ALL into their appropriate risk categories. In the model selection process, 103 clinical variables were used to train 4 classification algorithms, random forest (RF), decision tree, support vector machine (SVM), and linear regression, to distinguish relapses from nonrelapses in the 3 clinically predefined risk categories: standard-, intermediate-, and high-risk levels. While Pan et al [131] built a model to predict disease relapse, Hauser et al [144] studied the possibility of predicting CML prior to diagnosis using only CBC test results and ML algorithms such as XGBoost and LASSO algorithms on 1623 patients with a definitive CML status. The variables used in the study included laboratory CBC test results, patient demographic features such as their age and gender, and patient encounter information (the number of patient visits to outpatient clinics, etc.). A forward feature selection process was employed to measure the predictive performance of the most potential predictors. The data set was then divided into 7 subsets, in which time of diagnosis was set as a patient baseline, and the 6 remaining sets corresponded to the different time periods preceding the diagnosis test. Interestingly, variable selection yielded different features for inclusion in the models depending on the data collection interval.

The performances of the chosen classifiers in [131] were evaluated by a 10-fold cross validation in each of the 100 training sets. However, the suggested approach is considered insufficient for internal validation that requires at least 50 repeats [145]. By contrast, the chosen data set in [144] was divided into 2 distinct groups: train/validation and test groups. While the latter split-sample validation approach seems reasonable and justifiable to use in this case given the large sample size, potential drawbacks may arise, and several aspects still need attention throughout the application. For instance, as the sample split was performed fully at random, substantial patient imbalances might have occurred with respect to the distributions of predictors and the output. Moreover, 20% was used for model assessment leading to a potential biased evaluation of the model's performance [145].

Furthermore, it is well known that patients with leukemia often suffer from health issues due to frequent infections, which can lead to death if it is not detected early. Toward this end, Agius et al [55] investigated the risk of infection due to a weakness in the immune system or a cytotoxic treatment immediately after CLL diagnosis by developing the CLL Treatment-Infection Model (CLL-TIM). For each patient, the prediction point was set at 3 months after their diagnosis, and the target output was the 2-year infection risk or CLL treatment. After excluding 74 patients who died and 373 who initiated their treatment before the prediction point, the study cohort's final size corresponded to 3729 patients. In contrast to Hauser et al [144], Agius et al [55] employed stratified sampling to maintain class distributions and compensate for the 52% International Prognostic Index for CLL (CLL-IPI) missing variables, by dividing the data set into 65% training set and 17.5% for each of the test and internal validation sets. Using 7288 features resulting from a collection of variables from different sources, comprising baseline variables at the time of diagnosis including age, gender, etc.; routine laboratory tests; microbiology findings; pathology reports; and diagnosis codes for all patients, the CLL-TIM ensemble algorithm was composed of 28 ML models that could identify patients at a high risk of infection to increase their chances of survival. The model was then validated on both internal and independent external test cohorts, and exhibited interesting performances, surpassing the CLL-IPI.

Deep Learning–Based Models

Hassouneh et al [50] suggested the use of deep neural networks (DNNs) to predict survivability of patients with leukemia to boost the psychological state of patients and enable physicians to arrange the proper treatments for different cases. The final DNN structure encompassed 6 hidden layers, with 45 hidden neurons in each corresponding hidden layer, and a dropout activation with 25%. While Hassouneh et al [50] used patient records and attributes of patients with leukemia for modeling, Boldú et al [134] relied on a set of 731 blood smear images to predict initial patient diagnosis. For data set limitation purposes, 4 convolutional neural network (CNN) models were pretrained on a larger data set and used for the aforementioned task. Their respective performances were compared and the architecture achieving the best outcome was trained further. Next, these pretrained CNNs were fine-tuned to match the type of data that are fed into the model. Finally, the proposed ALNet model was able to distinguish, on a first level, between healthy and abnormal blood cell images. On a second level, it was able to identify whether the blasts were myeloid or lymphoid. After a thorough evaluation process using 5-fold cross validation and the hold-out (80%/20%) approach with 470 iterations, ALNet demonstrated interesting results and was chosen as the best architecture for modeling. One strength of this study is that the 5-fold cross validations were approximately balanced, and despite the random split, the data from the same patient smear were maintained within the same fold. Alternatively, the proposed DNN algorithm in [50] was trained and evaluated using 2 methods: 10-fold cross validation and ensemble method. Although the 10-fold cross-validation method is known to generate more stable results, the number of repetitions is quite crucial, which was absent in this study [145].

Overall, it is very clear that all the aforementioned algorithms succeeded in predicting the different aspects and repercussions of the disease. Nevertheless, no existing literature has yet examined the root cause of the malignancy by predicting the possibility of patient infection. Table 4 summarizes some studies in the prediction phase alongside their objectives, the data sets used in the studies, the methodologies followed, the performance of the models applied, their strengths and weaknesses, and the validation approach used, where available.

Table 4. Study analysis for journal publications on the prediction phase.

Reference	Objective, data set, and methodology	Performance and remarks
[131]	• Objective: ALL[a] relapse prediction • Data set: 336 newly diagnosed children with ALL • Methodology: Random forest algorithm	Performance: • Accuracy: 0.829 • AUC[b]: 0.902 Strengths: • Usage of 4 ML[c] algorithms and 104 features • Good model performance in all risk-level groups • Adoption of a special feature selection strategy: 100-fold Monte Carlo cross validation combined with 10-fold cross validation Limitations: • Data set imbalance (relapsed and nonrelapsed children) • Strong predictors were excluded from the variable set Validation: • 10-fold cross validation
[144]	• Objective: Prediction of patients with CML[d] and non-CML using complete blood count records • Data set: Complete blood count records of 1623 patients with a BCR-ABL1 test extracted from the US Veterans Health Administration • Methodology: XGBoost and LASSO	Performance: • AUC range: 0.87-0.96 at the time of diagnosis Strengths: • Use of 2 models • Use of 2 feature selection methods Limitations: • Imbalanced data set (predominant gender is male) • Nonstandard data collection process Validation: • Split sample validation (20% of the data for validation)
[1]	• Objective: Leukemia detection based on biomedical data • Data set: 401 leukemia datapoints from Z H Sikder Medical College and Hospital • Methodology: Decision tree	Performance: • Accuracy: 100% Strengths: • Use of 4 supervised ML algorithms Limitations: • Overfitting Validation: • 10-fold cross validation
[50]	• Objective: Prediction of leukemia survivability • Data set: 131,615 records and 133 attributes for patients with leukemia from the SEER[e] database • Methodology: Deep neural network model	Performance: • Accuracy: 74.85% Strengths: • Use of a DNN[f] ensemble method Limitations: • Many problems in the leukemia data set (redundant attributes, missing values, and unknown values) Validation: • 10-fold cross validation • Ensemble method

Reference	Objective, data set, and methodology	Performance and remarks
[96]	• Objective: Predictive identification of patients at risk during treatment • Data set: 737 samples of patients diagnosed with CLL[g] at Mayo Clinic • Methodology: logistic regression, support vector machine, gradient boosting machine, random forest	Performance: • ROC[h]-AUC: above 80% Strengths: • Binary classification outperforms survival analytic methods Limitations: • Lack of actionable information provided by the ML algorithms Validation: • 100 runs of 5-fold cross validation

[a]ALL: acute lymphoblastic leukemia.

[b]AUC: area under the curve.

[c]ML: machine learning.

[d]CML: chronic myeloid leukemia.

[e]SEER: Surveillance, Epidemiology, and End Results

[f]DNN: deep neural network.

[g]CLL: chronic lymphocytic leukemia.

[h]ROC: receiver operating characteristic

Hematology Screening

Machine Learning–Based Models

Initial leukemia screening and its efficient diagnosis require a deep and thorough image analysis process. As opposed to traditional manual screening, automated leukemia screening is a novel approach that minimizes human interaction and provides more accurate clinical information by using blood smear images to identify ALL automatically [17]. The automated screening process is considered challenging due to the leukocyte localization and region extraction phases, which are generally obtained via background removal and separation of surrounding blood components that might distort the overall detection process. For this reason, many studies have employed techniques such as principal component analysis as a filter to identify any features that do not bring any important information to the classification process [115] to enhance detection accuracy. Similarly, Chebouba et al [32] used a meta-heuristic stochastic local search technique to select the most important genes and proteins to be used in the RF-based classification of patients with AML.

Deep Learning–Based Models

CML consists of 3 sequential phases that change based on the patient's status and can progress to more severe phases if timely treatment is not provided. This makes CML phase identification very crucial, as different phases require separate treatments and medical regimens. In the chronic phase, less than 10% of the cells in both the blood and bone marrow are blasted. The severity and persistence of the aforementioned phase depend mainly on the consistency of the therapy followed. If the chronic stage is neglected and the patient does not receive timely and effective treatment, the condition can deteriorate to reach an accelerated phase where the blast count increases to around 10%-19%. Similarly, if the patient's condition declines with no appropriate medical intervention, the percentage of blasted WBCs doubles to reach around 20% or more. At this stage, the patient's state is considered uncontrollable, and the patient starts to exhibit symptoms such as fever, weakness, and weight loss [33]. As CNNs were proven to be efficient tools for accurate image recognition, Khosla and Ramesh [33] suggested using the latter to classify different CML images into their respective phases. Similarly, Togacar et al [71] used CNNs to separate WBC images into their 4 subclasses: eosinophil, lymphocyte, monocyte, and neutrophil. While the aforementioned CNNs are successful in identifying the most important features in images with no human supervision, they generally require large training data to achieve high performance, and their employment is regarded expensive in terms of both time and training. For this reason, while many studies tend to employ image augmentation to the existing data set to create larger samples by slightly changing the existing collected images [33], others implement the concept of transfer learning to overcome the training data shortage. The idea behind the latter is to leverage the power and knowledge of a pretrained model to apply it on a new similar task. For instance, Sahlol et al [3] proposed a novel approach consisting of a hybrid model and combined CNN feature extraction using the solid architecture of VGGNet that was pretrained on ImageNet, to separate malignant cells from benign ones. Similarly, Li et al [127] developed the globally optimized transfer deep-learning platform with multiple pretrained CNNs (GOTDP-MP-CNNs). This DL platform is composed of 17 CNNs able to classify pathologic images into human diffuse large B-cell lymphoma and non–diffuse large B-cell lymphoma. Table 5 summarizes some of the collected studies in the screening phase.

Table 5. Study analysis for journal papers in the screening phase.

Reference	Objective, data set, and methodology	Performance and remarks
[3]	• Objective: Classification of white blood cell leukemia • Data set: Acute Lymphoblastic Leukemia Image Database for Image Processing 1 and 2 • Methodology: A hybrid model (CNN[a] and SESSA[b])	Performance: • Accuracy: 99.2% • Sensitivity: 100% Strengths: • Powerful performance using CNN • Use of the salp swarm optimization method • Hybrid classification method • Use of transfer learning Limitations: • Small limited data set insufficient to train CNNs Validation: • 5-fold internal cross validation and 20% testing (external validation)
[61]	• Objective: Automated identification of acute lymphoblastic leukemia • Data set: Blood smear images obtained from the Department of Hematology at the University Hospital Ostrava • Methodology: support vector machine/artificial neural networks	Performance: • Accuracy: 98.19% Strengths: • High classification accuracy • Successful feature selection Limitations: • Extensive preprocessing is required • Lack of medical data sets • Inability to generalize the results and trends for lack of comparison with other methods Validation: • 10-fold cross validation repeated 10 times
[33]	• Objective: Classification of chronic myeloid leukemia phases • Data set: 500 pictures from Patliputra Medical College and Hospital, Dhanbad, and the blood journal repository • Methodology: CNN	Performance: • Accuracy: 97.8% Strengths: • Use of transfer learning Limitations: • Limited data set Validation: • Internal validation (14 left for testing)

[a]CNN: convolutional neural network.

[b]SESSA: statistically enhanced salp swarm algorithm.

Hematology Diagnosis

Machine Learning–Based Models

To address the challenge of manually detecting blasted cells, Dasariraju et al [54], Inbarani et al [66], Abedy et al [29], Jagadev and Virani [34], and Dharani and Hariprasath [31] used medical images of healthy and malignant samples to automatically identify the leukemic types and subtypes. While Dasariraju et al [54] applied an RF algorithm as an approach to differentiate between abnormal and healthy leukocytes, and classify immature leukocytes into their 4 subtypes, Inbarani et al [66] discussed the implementation of a novel sophisticated approach to identify ALL blast cells via the histogram-based soft covering rough K-means clustering (HSCRKM) segmentation algorithm. The latter is a hybrid-clustering technique that combines the strengths of both the soft covering rough set and the rough K-means clustering. Nevertheless, one main limitation of the HSCRKM segmentation technique is that it is not suitable for multiple color images because the latter increase the processing time due to an increase in the peak values of the histogram. To enhance image representation and prepare a clean input for the classification model, Jagadev and Virani [34] applied SVM on 220 blood smear images of healthy individuals and patients with leukemia to identify the 4 leukemic subtypes (AML, CML, CLL, and ALL) using both K-means clustering and hue, saturation, value color–based segmentation techniques. Similarly, Dasariraju et al [54] performed a set of morphological imaging modifications and preprocessing techniques to segment the nucleus and cytoplasm and overcome the difficulty of blood image detection. To reduce data

dimensionality, the model's speed, and processing time, the most relevant features were obtained using feature extraction techniques and the resulting output was passed on to the classifier [34,54,66]. Alternatively, Abedy et al [29] chose to employ the histogram of oriented gradients for feature extraction, the Gaussian filter for noise elimination, and the Sobel kernel for image filtering. Furthermore, Dey and Islam [49] adopted the principal component analysis technique followed by grid search for hyperparameter tuning, which significantly decreased the number of components from 7129 features to only 6 important parameters. This data transformation technique did not only reduce the computational time and made it faster, but it also helped giving better results. Alongside the usage of imaging techniques, many studies employed other techniques; for example, Moraes et al [133] suggested the usage of flow cytometry data for distinguishing leukemia/lymphoma, and Mahmood et al [143] directed their research to focus more on identifying the most discriminatory features for CLL using patient laboratory test results, demographic parameters, and training a Classification and Regression Trees model on 94 pediatric patients, which was evaluated using 10-fold cross validation. Moreover, both Dharani and Hariprasath [31] and Jagadev and Virani [34] used SVM to classify leukemia and its subtypes, while Paswan and Rathore [28] used K-nearest neighbors to separate blasted blood cells from normal ones and classify them further into either AML or ALL using a value of K=4. By contrast, Moraes et al [133] suggested the implementation of decision tree as an ML-based technique for distinguishing leukemia/lymphoma, where a binary classification between healthy and immature leukocytes was performed with an 80%/20% data split, followed by a subclassification of immature leukocytes into their respective 4 types using a 70%/30% split, and several combinations of hyperparameters were evaluated during a 5-fold cross validation. Conversely, Abedy et al [29] chose to employ logistic regression as a classifier to identify the shape of the leukemic cell from microscopic blood images, while Dey and Islam [49] conducted a study to detect patients' leukemia type based on their gene expression information using the RF algorithm and 2 other algorithms, XGBoost and artificial neural networks (ANNs), and Dasariraju et al [54] used RF to perform a subclassification of immature leukocytes into their respective 4 types.

Deep Learning–Based Models

With the objective of optimizing leukemia diagnosis, some studies made use of genetic features. For example, Rodrigues and Deusdado [63] suggested the application of a kernel logistic regression that classifies gene expression data using meta-learners to select the most relevant attributes before classification. The data set used for training comprised the 2 types of leukemia (ALL and AML) and a set of gene features. Pearson correlation and chi-square statistic were the 2 approaches used on meta-learners to assess attributes. Then, all models used 10-fold cross validation resulting in an identification of 12 common genes. Other studies, such as that by Al-Dulaimi et al [23], focused on current practices, techniques, and challenges in digital hematology detection of WBCs and their components (nuclei and cytoplasm) using hematologic microscopy images. In addition to analyzing the growing trends in computer-aided diagnosis applications, the review highlighted the main challenges associated with the use of CNNs in terms of both high computational time and costs to classify images and detect abnormalities. This gave rise to the use of transfer learning and enhancing optimization techniques, such as the bare bones particle swarm optimization algorithm used by Srisukkham et al [12] to extract the most informative features and enhance the classification accuracy of the lymphocytic cells into either normal or blasted. Likewise, Miyoshi et al [122] directed their work toward enhancing lymphoma diagnosis by classifying histopathological lymphoma images using a DL model. The aim of the study was to evaluate the performance of the suggested automated DL model and compare it with that of a traditional manual hematopathology detection procedure using CNNs. In this study, each test set comprised a total of 100 image patches, and the rest was randomly divided into 5 separate groups. During each repetition for 5 iterations, 1 group was kept for validation to evaluate the classifier performance for every epoch, while the other 4 groups were used for training. The number of epochs used in this study was 30. Similarly, Shafique and Tehsin [125] suggested the deployment of deep CNNs to detect the ALL type and its corresponding subtypes, while Zhao et al [60] proposed turning the captured raw multiparameter flow cytometry data into a 2D image by means of a self-organizing map (SOM) to analyze and classify them using the aforementioned algorithms. An SOM is a map of neurons that relies on unsupervised learning, in the sense that human intervention is not necessarily required. This model is used in many applications and has strong generalization abilities. The model was trained for 15 epochs, at a learning rate of 0.001 using an Adam optimizer function. The evaluation of classification accuracies used a 10% validation split, which was also employed for network architecture optimization. Then, the model performance was evaluated using the hold-out test set. In the same context, Sipes and Li [4] attempted fine-grained image classification for ALL diagnosis and compared the accuracies of CNNs and other models that used specific hand-selected features, while Vincent et al [37] proposed a leukemia classification that could be performed on 2 levels. The first process was applied to well-segmented nuclei extracted from 100 blood smear images. In this step, 90 samples were used for training and 10 were kept for validation. The cells were classified into normal and abnormal. The second step consisted of feeding 5 features extracted from abnormal images into a second classifier that split the images into ALL and AML types accordingly. Table 6 summarizes some of the studies in the diagnosis phase based on their objectives, data sets used, methodologies, and performance characteristics.

Table 6. Study analysis for journal publications on the diagnosis phase.

Reference	Objective, data set, and methodology	Performance and remarks
[60]	• Objective: Classification of mature B-cell neoplasm • Data set: 20,622 routine diagnostic samples from Munich Leukemia Laboratory • Methodology: CNN-SOM[a] transformation	Performance: • Accuracy: 95% Strengths: • Large data set • High accuracy Limitations: • Nonuniform distribution of misclassifications due to similarity in flow cytometric profiles Validation: • 10% validation split
[54]	• Objective: Detection of immature leukocytes and their classification into 4 types • Data set: Images extracted from a publicly available data set at The Cancer Imaging Archive • Methodology: Random forest algorithm	Performance: • Accuracy: 92.99% Strengths: • High precision results for each class Limitations: • High number of false positives leading to low precision and specificity Validation: • 5-fold cross validation
[49]	• Objective: Identification of the leukemia type based on patient genetic expression • Data set: A sample of 7129 genes that represent the genetic expressions of 72 people from Kaggle • Methodology: XGBoost, artificial neural networks, and random forest algorithm	Performance: • Random forest accuracy: 80.8% • XGBoost accuracy: 92.3% Strengths: • Use of principal component analysis for dimensionality reduction and faster computation • Use of grid search for the best hyperparameter selection Limitations: • Small data set (72 people) Validation: • Internal validation (65%/35% split)
[135]	• Objective: Classification of lymphocytic cells • Data set: The ALL-IDB2 Database • Methodology: bare bones particle swarm optimization–based feature optimization	Performance: • Accuracy: 94.94%-96.25% Strengths: • A good performance on capturing prognostic chronic myeloid leukemia markers by the model Limitations: • Challenge of capturing relationships between data types with no information loss in clinical clustering Validation: • Validation on an external independent clinical trial

Reference	Objective, data set, and methodology	Performance and remarks
[34]	• Objective: Detection of leukemia and its types • Data set: 220 blood smear images from healthy individuals and patients with leukemia • Methodology: support vector machine	Performance: • Accuracy: Above 80% Strengths: • Use of 3 segmentation methods • Broader range of leukemia classification (types and subtypes) Limitations: • Costly method based on imaging data Validation: • Internal validation (train test split)
[122]	• Objective: Automated detection of malignant lymphoma • Data set: Prepared histopathologic images (388 sections, 259 diffuse large B-cell lymphomas, 89 follicular lymphomas, and 40 reactive lymphoid hyperplasia) • Methodology: Deep neural network classifier	Performance: • Accuracy: 97% Strengths: • High accuracy outperforming 7 pathologists • Model ensemble comprising 3 classifiers Limitations: • Classifier requires a manual annotation • Model not able to classify all the subtypes Validation: • K-fold cross validation repeated 5 times
[37]	• Objective: Multiclassification of leukemia • Data set: 100 blood smear images • Methodology: Neural network classifiers	Performance: • Accuracy: 97.7% Strengths: • Two-step neural network classifier Limitations: • Limited data set (100 blood smear images) Validation: • Internal validation (90 images used for training and 10 kept for validation)
[133]	• Objective: Leukemia and lymphoma diagnosis • Data set: 283 blood and bone marrow sample images from patients with leukemia and lymphoma • Methodology: Decision tree	Performance: • Correctness: 95% Strengths: • Application of the LASSO algorithm for regularization • Model robustness and strength against false negatives Limitations: • Complexity of the decision tree and the risk of overfitting through the production of too large trees Validation: • 30-fold cross validation
[66]	• Objective: Leukemia image segmentation • Data set: The Acute Lymphoblastic Leukemia Image Database • Methodology: HSCRKM[b]/particle swarm optimization/K-means	

Reference	Objective, data set, and methodology	Performance and remarks
		Performance: • Accuracy: 80% and above Strengths: • Use of 7 machine learning methods • Application of soft covering rough approximation Limitations: • Suitable for medical images only • Application on multiple color images increases the processing time Validation: • Different train/test sizes were used for model evaluation
[144]	• Objective: Determining the most predictive features for acute lymphoblastic leukemia identification • Data set: 94 pediatric patient samples collected from the Department of Hematology and Oncology, Children Hospital and Institute of Child Health, Lahore • Methodology: Random forest, boosting machine, C5.0 decision tree, and classification and regression trees	Performance: • Accuracy: 87.4% Strengths: • High accuracy • Balanced data set Limitations: • Small-scale study • Few machine learning models • Socioeconomic risk factors not selected automatically Validation: • Internal validation (train/validation data) • 10-fold cross validation
[31]	• Objective: Leukemia diagnosis and its subtypes • Data set: 200 blood smear images extracted from Vidyalankar Institute of Technology, Mumbai and online databases • Methodology: support vector machine	Performance: • Accuracy: 97.8% Strengths: • Good detection accuracy • Thorough image segmentation process Limitations: • Challenging detection process due to the irregularity of the cancer cell's shape and nucleus • Use of only support vector machine for classification

[a]CNN-SOM: convolutional neural network-self-organizing map.

[b]HSCRKM: histogram-based soft covering rough K-means clustering.

Hematology Treatment

Machine Learning–Based Models

Complete remission (CR) refers to the disappearance of all signs and symptoms of an illness [51]. However, a significant proportion of patients report disease relapse after therapy and complete disease recovery. In this regard, Gal et al [75] proposed an ML technique that uses gene expressions to predict the likelihood of CR in patients with AML who previously received therapy. The 473 collected samples were divided into training and testing and were fed into the 3 classifiers for feature selection using a 5-fold cross validation. To select the most significant genes that clearly mark the difference between the state of CR and non-CR, a statistical *t* test was performed in each fold for each method. For further gene feature selection and performance enhancement, the results were compared with those of 3 algorithms: randomized LASSO, recursive feature elimination, and hill climbing. It was proven that cancers that appear pathologically identical do not necessarily exhibit the exact response to similar drugs. To overcome the challenge of high demands for personalized or patient-specific medicine, Lee et al [9] proposed using a gene-expression profile and in vitro drug sensitivity data to spot molecular markers that explain this patient-specific drug response. The data set used comprised 160 chemotherapy drugs and inhibitors for 30 patients with AML. The gene-drug association was identified using the MERGE algorithm, which utilized gene characteristics such as a novel mutation, expression hubness, known regulator, genomic copy number variation, and methylation. The model testing was done in 2 different ways: the first approach used 2 batches containing 12 patient samples for training and 12 different samples for validation, respectively. The second approach used a leave-one-out cross validation to test the predicted drug sensitivity for 30 patient samples. The latter cross-validation

method is regarded as reasonable because it prevents the high computational training cost and time owing to the small sample size and is much less biased than using a single test set because the process fits the data set consisting of n-1 observations repeatedly [145]. Although having a limited data size permitted the obtention of efficient validation results, identifying gene-drug associations is deemed to be a challenging process in that case. However, the proposed algorithm was successful to prioritize genes based on the multidimensional data on their potential to drive cancer. Consequently, upon comparison with other alternate methods (ElasticNet, multitask learning, Pearson P value, and Spearman P value), MERGE exhibited the best gene-drug association result.

Deep Learning–Based Models

In their review and analysis of current AI applications for hematologic disorders' treatment, Muhsen et al [24] presented a study that was performed with a set of patients who underwent allogeneic hematopoietic cell transplantation to predict the development of acute graft-versus-host disease using ANNs. After comparing the performance acquired by ANNs and the results achieved by logistic regression, ANNs were found to predict the presence of graft-versus-host disease significantly better. However, among the limitations remaining to help reach optimal ML results is the limited data input from patients that could be further enlarged to include biologic and genetic factors, for instance. Moreover, the employment of several sampling techniques such as random oversampling, synthetic oversampling, and remote under sampling could help improve the ML models' accuracies in predicting treatment-related mortality in allogeneic hematopoietic cell transplantation.

Similarly, Lyu et al [42] used ANNs to classify the progress and change in gene expressions and peripheral blood mononuclear cells before treatment and after starting therapy. Table 7 shows some extracted literature focusing on the hematology treatment phase.

Table 7. Study analysis for journal publications on the treatment phase.

Reference	Objective, data set, and methodology	Performance and remarks
[9]	• Objective: Digital analysis of blood smears and preclassification of cells • Data set: Images of blood smears from a hematologic laboratory • Methodology: MERGE algorithm	Performance: • Accuracy: 90% Strengths: • Introduction of a new computational and statistical method to determine gene markers Limitations: • Small data set comprising only 30 patients with acute myeloid leukemia Validation: • Leave-one-out cross validation
[75]	• Objective: Prediction of complete remission of acute myeloid leukemia • Data set: 473 bone marrow samples from the Children's Oncology Group • Methodology: K-nearest neighbor, support vector machine, and hill climbing	Performance: • Area under the curve: 0.84 Strengths: • Use of 3 feature selection algorithms: randomized LASSO, recursive feature elimination, and hill climbing • Use of 3 classifiers: support vector machine, random forest, and K-nearest neighbor Limitations: • Small data set Validation: • 100 iterations of a 5-fold cross validation
[81]	• Objective: Identify the right patterns to improve risk stratification of patient with CLLs[a] • Data set: (1) the first cohort comprised CLL cells of 196 individuals; the second cohort comprised CLL cells of 98 individuals including their clinical data and RNA-seq • Methodology: (1) EM algorithm and the Gaussian mixture models; (2) Boosted tree ensemble method	Performance: • Precision: 90% Strengths: • High accuracy and precision Limitations: • Large data set and 5-year monitoring is required Validation: • External validation on an independent cohort

[a]CLL: chronic lymphocytic leukemia.

Conclusions and Future Research Directions

Early diagnosis and prediction of hematologic malignancies can immensely reduce mortality rates and can improve patient survival rates. Nevertheless, the nature of data on medical treatment is complex and requires an in-depth analysis to extract the important explicative features and hidden data patterns. The only way to handle enormous sets is through the use of AI. The challenge that faces AI applications, however, is the limitation in data availability, which can be overcome by means of data augmentation techniques, regularization, and transfer learning. This review of the literature highlights the most recent applications of both DL and ML in the field of blood cancer management for every hematologic pathway stage and malignancy type. Based on the reviewed articles, ML techniques have been widely used, in comparison with DL methods, as the latter are relatively newer and require larger data sets than ML, which is considered a constraint in the medical field. In some studies, ML techniques performed better than DL methods and vice versa, depending on the application and nature of data used. Moreover, screening and diagnosis are challenging tasks, as hematologic cancers are difficult to identify during their initial stages. Therefore, many studies in the field investigated the aforementioned stages alongside prediction, while less attention has been paid to the treatment stage. The latter is critical and requires further analysis and study, as repercussions and relapses may arise due to cancer treatment, namely, chemotherapy, which requires a risk evaluation and future mitigation plans. Furthermore, some malignancies appeared to be more addressed than others, mainly acute myeloid/lymphoblastic leukemias that have gotten the most attention in the last few years followed by lymphoma, due to their fast development and aggressivity. Conversely, there was less emphasis and minor existing literature tackling the chronic types of leukemia. This can be due to the slow-growing pattern of the aforementioned, and a lack of sudden symptom exhibition until very late stages, which makes the latter's monitoring quite complex. Overall, a lack of detection accuracy can have a significant impact on the patient's journey to treatment because it can delay diagnosis and affect the efficiency of therapies. Therefore, predictive models that can recognize disease patterns and common symptoms in hematologic malignancies based on medical patient records are essential to forecast the risk of infection and avoid late-stage diagnoses. Thus far, many studies employed several techniques to predict hematologic malignancies' diagnosis through either medical image recognition, flow cytometry, or genetic expressions. However, no study in the literature has ever made use of patient CBC test results alone for blood disorder prediction or detection purposes. As the latter is generally regarded as the first diagnostic routine for hematologists to confirm leukemia diagnosis, it can be an efficient medium to potentially investigate in the future.

Acknowledgments

This article was made possible by the National Priorities Research Program-Standard (NPRP-S) Twelfth (12th) Cycle grant (NPRP12S-0219-190108) from the Qatar National Research Fund (a member of Qatar Foundation). The findings herein reflect the work, and are solely the responsibility, of the author[s].

Conflicts of Interest

None declared.

References

1. Das M, Kumar Saha P. A Comprehensive Study of Machine Learning algorithms for Predicting Leukemia Based on Biomedical Data. 2020 Dec 21 Presented at: International Conference on Innovation in Engineering and Technology; December 23-24, 2019; Dhaka, Bangladesh p. 1-5. [doi: 10.1109/ICIET48527.2019.9290544]
2. Acharya V, Ravi V, Pham TD, Chakraborty C. Peripheral Blood Smear Analysis Using Automated Computer-Aided Diagnosis System to Identify Acute Myeloid Leukemia. IEEE Trans. Eng. Manage 2021:1-14. [doi: 10.1109/tem.2021.3103549]
3. Sahlol AT, Kollmannsberger P, Ewees AA. Efficient Classification of White Blood Cell Leukemia with Improved Swarm Optimization of Deep Features. Sci Rep 2020 Feb 13;10(1):2536 [FREE Full text] [doi: 10.1038/s41598-020-59215-9] [Medline: 32054876]
4. Sipes R, Li D. Using Convolutional Neural Networks for Automated Fine Grained Image Classification of Acute Lymphoblastic Leukemia. 2019 May Presented at: 2018 3rd International Conference on Computational Intelligence and Applications (ICCIA); July 28-30, 2018; Hong Kong, China p. 157-161. [doi: 10.1109/ICCIA.2018.00036]
5. Ratley A, Minj J, Patre P. Leukemia Disease Detection and Classification Using Machine Learning Approaches: A Review. 2020 Presented at: 2020 First International Conference on Power, Control and Computing Technologies (ICPC2T); January 3-5, 2020; Raipur, Chhattisgarh, India p. 161-165. [doi: 10.1109/icpc2t48082.2020.9071471]
6. Schenone D, Dominietto A, Campi C, Frassoni F, Cea M, Aquino S, et al. Radiomics and Artificial Intelligence for Outcome Prediction in Multiple Myeloma Patients Undergoing Autologous Transplantation: A Feasibility Study with CT Data. Diagnostics (Basel) 2021 Sep 24;11(10):1759 [FREE Full text] [doi: 10.3390/diagnostics11101759] [Medline: 34679456]
7. Shouval R, Labopin M, Bondi O, Shamay H, Shimoni A, Ciceri F. Prediction of allogeneic hematopoietic stem cell transplantation (Allo-HSCT) related mortality in acute leukemia: Generation of a machine learning-based model using the

data set of the Acute Leukemia Working Party (ALWP) of the EBMT. Blood 2013;122(21):409. [doi: 10.1182/blood.v122.21.409.409]
8. Renuka T, Surekha B. Acute-lymphoblastic leukemia detection through deep transfer learning approach of Neural Network. 2021 Mar 19 Presented at: First Doctoral Symposium on Natural Computing Research; August 8, 2020; Pune, Maharashtra, India p. 163-170. [doi: 10.1007/978-981-33-4073-2_17]
9. Lee S, Celik S, Logsdon BA, Lundberg SM, Martins TJ, Oehler VG, et al. A machine learning approach to integrate big data for precision medicine in acute myeloid leukemia. Nat Commun 2018 Jan 03;9(1):42 [FREE Full text] [doi: 10.1038/s41467-017-02465-5] [Medline: 29298978]
10. Avalos M, Touchais H, Henríquez-Henríquez M. Optimising criteria for Manual Smear Review following Automated Blood Count Analysis: A machine learning approach. Advances in Intelligent Systems and Computing 2021 Apr 10;91:380-391. [doi: 10.1007/978-3-030-73603-3_35]
11. Darane B, Rajput P, Sondagar Y, Koshy R. Recognizing Presence of Hematological Disease using Deep Learning. 2020 Mar 12 Presented at: 2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC); December 12-14, 2019; Palladam, Tamil Nadu, India p. 310-315. [doi: 10.1109/I-SMAC47947.2019.9032639]
12. Srisukkham W, Zhang L, Neoh SC, Todryk S, Lim CP. Intelligent leukaemia diagnosis with bare-bones PSO based feature optimization. Applied Soft Computing 2017 Jul;56:405-419. [doi: 10.1016/j.asoc.2017.03.024]
13. Radakovich N, Cortese M, Nazha A. Acute myeloid leukemia and artificial intelligence, algorithms and new scores. Best Pract Res Clin Haematol 2020 Sep;33(3):101192 [FREE Full text] [doi: 10.1016/j.beha.2020.101192] [Medline: 33038981]
14. Rodellar J, Alférez S, Acevedo A, Molina A, Merino A. Image processing and machine learning in the morphological analysis of blood cells. Int J Lab Hem 2018 May 09;40:46-53. [doi: 10.1111/ijlh.12818]
15. Salah HT, Muhsen IN, Salama ME, Owaidah T, Hashmi SK. Machine learning applications in the diagnosis of leukemia: Current trends and future directions. Int J Lab Hematol 2019 Dec 09;41(6):717-725. [doi: 10.1111/ijlh.13089] [Medline: 31498973]
16. Kratz A, Lee S, Zini G, Riedl JA, Hur M, Machin S, International Council for Standardization in Haematology. Digital morphology analyzers in hematology: ICSH review and recommendations. Int J Lab Hematol 2019 Aug 02;41(4):437-447. [doi: 10.1111/ijlh.13042] [Medline: 31046197]
17. Benard B, Gentles AJ, Köhnke T, Majeti R, Thomas D. Data mining for mutation-specific targets in acute myeloid leukemia. Leukemia 2019 Apr 6;33(4):826-843 [FREE Full text] [doi: 10.1038/s41375-019-0387-y] [Medline: 30728456]
18. Banjar H, Adelson D, Brown F, Chaudhri N. Intelligent Techniques Using Molecular Data Analysis in Leukaemia: An Opportunity for Personalized Medicine Support System. BioMed Research International 2017 Jul 25;2017:1-21. [doi: 10.1155/2017/3587309]
19. El Achi HE, Khoury JD. Artificial Intelligence and Digital Microscopy Applications in Diagnostic Hematopathology. Cancers (Basel) 2020 Mar 26;12(4):797 [FREE Full text] [doi: 10.3390/cancers12040797] [Medline: 32224980]
20. Eckardt J, Bornhäuser M, Wendt K, Middeke J. Application of machine learning in the management of acute myeloid leukemia: current practice and future prospects. Blood Adv 2020 Dec 08;4(23):6077-6085 [FREE Full text] [doi: 10.1182/bloodadvances.2020002997] [Medline: 33290546]
21. Umamaheswari D, Geetha S. Review on image segmentation techniques incorporated with machine learning in the scrutinization of leukemic Microscopic Stained Blood Smear Images. 2019 Jan 02 Presented at: International Conference on ISMAC in Computational Vision and Bio-Engineering (ISMAC-CVB); May 2018; Palladam, Tamil Nadu, India p. 1773-1791. [doi: 10.1007/978-3-030-00665-5_163]
22. Khan S, Sajjad M, Hussain T, Ullah A, Imran AS. A Review on Traditional Machine Learning and Deep Learning Models for WBCs Classification in Blood Smear Images. IEEE Access 2021;9:10657-10673. [doi: 10.1109/access.2020.3048172]
23. AL-Dulaimi K, Banks J, Nugyen K, Al-Sabaawi A, Tomeo-Reyes I, Chandran V. Segmentation of White Blood Cell, Nucleus and Cytoplasm in Digital Haematology Microscope Images: A Review–Challenges, Current and Future Potential Techniques. IEEE Rev. Biomed. Eng 2021;14:290-306. [doi: 10.1109/rbme.2020.3004639]
24. Muhsen IN, ElHassan T, Hashmi SK. Artificial Intelligence Approaches in Hematopoietic Cell Transplantation: A Review of the Current Status and Future Directions. Turk J Haematol 2018 Aug 03;35(3):152-157 [FREE Full text] [doi: 10.4274/tjh.2018.0123] [Medline: 29880463]
25. Alagu S, Bagan K. Computer Assisted Classification Framework for detection of acute myeloid leukemia in Peripheral Blood Smear Images. In: Innovations in Computational Intelligence and Computer Vision. Singapore City, Singapore: Springer; Sep 22, 2020:403-410.
26. Akter HM, Islam SM, Muntasir I, Muzahidul IA, Islam S, Ahmed A. Leukemia detection mechanism through microscopic image and ML Techniques. 2020 Dec 22 Presented at: 2020 IEEE Region 10 Conference (TENCON); November 16-19, 2020; Osaka, Japan. [doi: 10.1109/tencon50793.2020.9293925]
27. Mandal S, Daivajna V, Rajagopalan V. Machine Learning based System for Automatic Detection of Leukemia Cancer Cell. 2020 Mar 12 Presented at: 2019 IEEE 16th India Council International Conference (INDICON); December 13-15, 2019; Rajkot, Gujarat, India p. 1-4. [doi: 10.1109/indicon47234.2019.9029034]
28. Paswan S, Rathore Y. Recognition and Arrangement of Blood Cancer from Microscopic Cell Pictures Utilizing Support Vector Machine K-Nearest Neighbor and Deep Learning. 2019 Mar 18 Presented at: 2018 International Conference on

Communication, Computing and Internet of Things (IC3IoT); February 15-17, 2018; Chennai, Tamil Nadu, India p. 525-530. [doi: 10.1109/ic3iot.2018.8668141]
29. Abedy H, Ahmed F, Bhuiyan M, Islam M, Ali NY, Shamsujjoha M. Leukemia Prediction from Microscopic Images of Human Blood Cell Using HOG Feature Descriptor and Logistic Regression. 2019 Jan 17 Presented at: 2018 16th International Conference on ICT and Knowledge Engineering (ICT&KE); November 21-23, 2018; Bangkok, Thailand p. 1-6. [doi: 10.1109/ictke.2018.8612303]
30. Rajpurohit S, Patil S, Choudhary N, Gavasane S, Kosamkar P. Identification of Acute Lymphoblastic Leukemia in Microscopic Blood Image Using Image Processing and Machine Learning Algorithms. 2018 Dec 03 Presented at: 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI); September 19-22, 2018; Bengaluru, Karnataka, India p. 2359-2363. [doi: 10.1109/icacci.2018.8554576]
31. Dharani T, Hariprasath S. Diagnosis of Leukemia and its types Using Digital Image Processing Techniques. 2019 May 30 Presented at: 2018 3rd International Conference on Communication and Electronics Systems (ICCES); October 15-16, 2018; Coimbatore, Tamil Nadu, India p. 275-279. [doi: 10.1109/cesys.2018.8724075]
32. Chebouba L, Boughaci D, Guziolowski C. Proteomics Versus Clinical Data and Stochastic Local Search Based Feature Selection for Acute Myeloid Leukemia Patients' Classification. J Med Syst 2018 Jun 04;42(7):129. [doi: 10.1007/s10916-018-0972-z] [Medline: 29869179]
33. Khosla E, Ramesh D. Phase classification of chronic myeloid leukemia using convolution neural networks. 2018 Jun 21 Presented at: 2018 4th International Conference on Recent Advances in Information Technology (RAIT); March 15-17, 2018; Dhanbad, Jharkhand, India p. 1-6. [doi: 10.1109/rait.2018.8389068]
34. Jagadev P, Virani HG. Detection of leukemia and its types using image processing and machine learning. 2018 Feb 22 Presented at: 2017 International Conference on Trends in Electronics and Informatics (ICEI); May 11-12, 2017; Tirunelveli, Tamil Nadu, India p. 522-526. [doi: 10.1109/icoei.2017.8300983]
35. Bahra G, Wiese L. Classifying leukemia and gout patients with Neural Networks. In: International Conference on Database and Expert Systems Applications. Cham, Switzerland: Springer; 2018:150-160.
36. Yu W, Chang J, Yang C, Zhang L, Shen H, Xia Y, et al. Automatic classification of leukocytes using deep neural network. 2018 Jan 11 Presented at: 2017 IEEE 12th International Conference on ASIC (ASICON); October 25-28, 2017; Guiyang, China p. 1041-1044. [doi: 10.1109/asicon.2017.8252657]
37. Vincent I, Kwon KR, Lee SH, Moon KS. Acute lymphoid leukemia classification using two-step neural network classifier. 2015 May 11 Presented at: 2015 21st Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV); January 28-30, 2015; Mokpo, South Korea p. 1-4. [doi: 10.1109/fcv.2015.7103739]
38. Chen R, Garapati S, Wu D, Ko S, Falk S, Dierov D. Machine learning based predictive model of 5-year survival in multiple myeloma autologous transplant patients. Blood. (Supplement_1) 2019;134:2156. [doi: 10.1182/blood-2019-129432]
39. Madanan M. Designing an Artificial Intelligence Model using Machine Learning Algorithms and Applying it to Hematology for the Detection and Classification of Various Stages of Blood Cancer. 2020 Jul Presented at: International Conference on Innovative Technical Advances in Disaster Management (ICITADM); July 15-16, 2020; Kalady, Kerala, India.
40. Hossain M, Sabik M, Rahman M, Sakiba S, Muzahidul IA, Shatabda S. An effective leukemia prediction technique using supervised machine learning classification algorithm. Singapore City, Singapore: Springer; 2020 Dec 17 Presented at: International Conference on Trends in Computational and Cognitive Engineering; December 17-18, 2020; Dhaka, Bangladesh p. 219-229. [doi: 10.1007/978-981-33-4673-4_19]
41. Licandro R, Reiter M, Diem M, Dworzak M, Schumich A, Kampel M. Application of machine learning for automatic MRD Assessment in Paediatric Acute Myeloid leukaemia. 2018 Presented at: 7th International Conference on Pattern Recognition Applications and Methods; January 16-18, 2018; Madeira, Portugal. [doi: 10.5220/0006595804010408]
42. Lyu M, Radenkovic M, Keskin D, Brusic V. Classification of single cell types during leukemia therapy using artificial neural networks. 2021 Jan 13 Presented at: 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); December 16-19, 2020; Seoul, South Korea p. 1258-1261. [doi: 10.1109/bibm49941.2020.9313319]
43. Patil BS, Mishra S, Junnarkar A. Leukemia diagnosis based on machine learning algorithms. 2021 Jan 01 Presented at: 2020 IEEE International Conference for Innovation in Technology (INOCON). 2020; November 6-8, 2020; Bangluru, Karnataka, India. [doi: 10.1109/inocon50539.2020.9298321]
44. Malkawi A, Al-Assi R, Salameh T, Sheyab B, Alquran H, Alqudah A. White Blood Cells Classification Using Convolutional Neural Network Hybrid System. 2020 Nov 25 Presented at: 2020 IEEE 5th Middle East and Africa Conference on Biomedical Engineering (MECBME); October 27-29, 2020; Amman, Jordan p. 1-5. [doi: 10.1109/mecbme47393.2020.9265154]
45. Raina S, Khandelwal A, Gupta S, Leekha A. Blood cells detection using faster-RCNN. 2020 Nov 02 Presented at: 2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON); October 2-4, 2020; Greater Noida, Uttar Pradesh, India. [doi: 10.1109/gucon48875.2020.9231134]
46. Singh G, Bathla G, Kaur S. A review to detect leukemia cancer in medical images. 2017 Jan 16 Presented at: International Conference on Computing, Communication and Automation (ICCCA); May 5-6, 2017; Greater Noida, NCR, India. [doi: 10.1109/ccaa.2016.7813896]

47. Tambe R, Mahajan S, Shah U, Agrawal M, Garware B. Towards designing an automated classification of lymphoma subtypes using Deep Neural Networks. 2019 Jan 03 Presented at: ACM India Joint International Conference on Data Science and Management of Data; January 3-5, 2019; Kolkata, West Bengal, India. [doi: 10.1145/3297001.3297019]
48. Bhuiyan MNQ, Rahut SK, Tanvir RA, Ripon S. Automatic Acute Lymphoblastic Leukemia Detection and Comparative Analysis from Images. 2019 Presented at: International Conference on Control, Decision and Information Technologies (CoDIT); April 23-26, 2019; Paris, France p. 1144-1149. [doi: 10.1109/codit.2019.8820299]
49. Dey U, Islam M. Genetic Expression Analysis To Detect Type Of Leukemia Using Machine Learning. 2019 Dec 19 Presented at: 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT); May 3-5, 2019; Dhaka, Bangladesh p. 1-6. [doi: 10.1109/icasert.2019.8934628]
50. Hassouneh N, Alnemer L, Alsakran J, Rodan A. Predicting Survivability in Leukemia Patients using Deep Learning. 2020 Presented at: 2019 Sixth HCT Information Technology Trends (ITT) Conference; 2019; Ras Al Khaimah, United Arab Emirates. [doi: 10.1109/itt48889.2019.9075067]
51. Jayashanka R, Wijesinghe C, Weerasinghe A, Pieris D. Machine Learning Approach to Predict the Survival Time of Childhood Acute Lymphoblastic Leukemia Patients. 2018 Presented at: 2018 18th International Conference on Advances in ICT for Emerging Regions (ICTer); September 26-29, 2018; Colombo, Sri Lanka p. 426-432. [doi: 10.1109/icter.2018.8615472]
52. Shurrab A, Maghari A. Blood diseases detection using data mining techniques. 2017 Presented at: 2017 8th International Conference on Information Technology (ICIT); May 17-18, 2017; Amman, Jordan p. 625-631. [doi: 10.1109/icitech.2017.8079917]
53. Chulián S, Martínez-Rubio Á, Pérez-García VM, Rosa M, Blázquez Goñi C, Rodríguez Gutiérrez JF, et al. High-Dimensional Analysis of Single-Cell Flow Cytometry Data Predicts Relapse in Childhood Acute Lymphoblastic Leukaemia. Cancers (Basel) 2020 Dec 23;13(1):17 [FREE Full text] [doi: 10.3390/cancers13010017] [Medline: 33374500]
54. Dasariraju S, Huo M, McCalla S. Detection and Classification of Immature Leukocytes for Diagnosis of Acute Myeloid Leukemia Using Random Forest Algorithm. Bioengineering (Basel) 2020 Oct 01;7(4):120 [FREE Full text] [doi: 10.3390/bioengineering7040120] [Medline: 33019619]
55. Agius R, Brieghel C, Andersen MA, Pearson AT, Ledergerber B, Cozzi-Lepri A, et al. Machine learning can identify newly diagnosed patients with CLL at high risk of infection. Nat Commun 2020 Jan 17;11(1):363 [FREE Full text] [doi: 10.1038/s41467-019-14225-8] [Medline: 31953409]
56. Gimenez N, Tripathi R, Giró A, Rosich L, López-Guerra M, López-Oreja I, et al. Systems biology drug screening identifies statins as enhancers of current therapies in chronic lymphocytic leukemia. Sci Rep 2020 Dec 17;10(1):22153 [FREE Full text] [doi: 10.1038/s41598-020-78315-0] [Medline: 33335123]
57. Sun J, Wang L, Liu Q, Tárnok A, Su X. Deep learning-based light scattering microfluidic cytometry for label-free acute lymphocytic leukemia classification. Biomed Opt Express 2020 Nov 01;11(11):6674-6686 [FREE Full text] [doi: 10.1364/BOE.405557] [Medline: 33282516]
58. Nagata Y, Zhao R, Awada H, Kerr C, Mirzaev I, Kongkiatkamon S, et al. Machine learning demonstrates that somatic mutations imprint invariant morphologic features in myelodysplastic syndromes. Blood 2020 Nov 12;136(20):2249-2262 [FREE Full text] [doi: 10.1182/blood.2020005488] [Medline: 32961553]
59. Kashef AA, Khatibi T, Mehrvar A. Prediction of Cranial Radiotherapy Treatment in Pediatric Acute Lymphoblastic Leukemia Patients Using Machine Learning: A Case Study at MAHAK Hospital. Asian Pac J Cancer Prev 2020 Nov 01;21(11):3211-3219 [FREE Full text] [doi: 10.31557/APJCP.2020.21.11.3211] [Medline: 33247677]
60. Zhao M, Mallesh N, Höllein A, Schabath R, Haferlach C, Haferlach T, et al. Hematologist-Level Classification of Mature B-Cell Neoplasm Using Deep Learning on Multiparameter Flow Cytometry Data. Cytometry A 2020 Oct 04;97(10):1073-1080 [FREE Full text] [doi: 10.1002/cyto.a.24159] [Medline: 32519455]
61. Bodzas A, Kodytek P, Zidek J. Automated Detection of Acute Lymphoblastic Leukemia From Microscopic Images Based on Human Visual Perception. Front Bioeng Biotechnol 2020;8:1005 [FREE Full text] [doi: 10.3389/fbioe.2020.01005] [Medline: 32984283]
62. Gaidano V, Tenace V, Santoro N, Varvello S, Cignetti A, Prato G, et al. A Clinically Applicable Approach to the Classification of B-Cell Non-Hodgkin Lymphomas with Flow Cytometry and Machine Learning. Cancers (Basel) 2020 Jun 24;12(6):1684 [FREE Full text] [doi: 10.3390/cancers12061684] [Medline: 32599959]
63. Rodrigues V, Deusdado S. Metalearning approach for leukemia informative genes prioritization. J Integr Bioinform 2020 May 08;17(1):7 [FREE Full text] [doi: 10.1515/jib-2019-0069] [Medline: 32383690]
64. Haider RZ, Ujjan IU, Shamsi TS. Cell Population Data-Driven Acute Promyelocytic Leukemia Flagging Through Artificial Neural Network Predictive Modeling. Transl Oncol 2020 Jan;13(1):11-16 [FREE Full text] [doi: 10.1016/j.tranon.2019.09.009] [Medline: 31733590]
65. Bibi N, Sikandar M, Ud Din I, Almogren A, Ali S. IoMT-Based Automated Detection and Classification of Leukemia Using Deep Learning. J Healthc Eng 2020;2020:6648574 [FREE Full text] [doi: 10.1155/2020/6648574] [Medline: 33343851]
66. Inbarani H, Azar A, Jothi G. Leukemia Image Segmentation Using a Hybrid Histogram-Based Soft Covering Rough K-Means Clustering Algorithm. Electronics 2020 Jan 19;9(1):188. [doi: 10.3390/electronics9010188]

67. Haferlach T. Advancing leukemia diagnostics: Role of Next Generation Sequencing (NGS) in acute myeloid leukemia. Hematol Rep 2020 Sep 21;12(Suppl 1):8957 [FREE Full text] [doi: 10.4081/hr.2020.8957] [Medline: 33042506]
68. Kumar D, Jain N, Khurana A, Mittal S, Satapathy S, Senkerik R, et al. Automatic Detection of White Blood Cancer From Bone Marrow Microscopic Images Using Convolutional Neural Networks. IEEE Access 2020;8:142521-142531. [doi: 10.1109/ACCESS.2020.3012292]
69. Khilji IQ, Saha K, Amin J, Iqbal M. Application of Homomorphic Encryption on Neural Network in Prediction of Acute Lymphoid Leukemia. IJACSA 2020;11(6):e1. [doi: 10.14569/IJACSA.2020.0110646]
70. Vasighizaker A, Sharma A, Dehzangi A. A novel one-class classification approach to accurately predict disease-gene association in acute myeloid leukemia cancer. PLoS One 2019;14(12):e0226115 [FREE Full text] [doi: 10.1371/journal.pone.0226115] [Medline: 31825992]
71. Togacar M, Ergen B, Sertkaya ME. Subclass Separation of White Blood Cell Images Using Convolutional Neural Network Models. Elektron Elektrotech 2019 Oct 06;25(5):63-68. [doi: 10.5755/j01.eie.25.5.24358]
72. Ahmed N, Yigit A, Isik Z, Alpkocak A. Identification of Leukemia Subtypes from Microscopic Images Using Convolutional Neural Network. Diagnostics (Basel) 2019 Aug 25;9(3):104 [FREE Full text] [doi: 10.3390/diagnostics9030104] [Medline: 31450720]
73. Fuse K, Uemura S, Tamura S, Suwabe T, Katagiri T, Tanaka T, et al. Patient-based prediction algorithm of relapse after allo-HSCT for acute Leukemia and its usefulness in the decision-making process using a machine learning approach. Cancer Med 2019 Sep;8(11):5058-5067 [FREE Full text] [doi: 10.1002/cam4.2401] [Medline: 31305031]
74. Reiter M, Diem M, Schumich A, Maurer-Granofszky M, Karawajew L, Rossi JG, International Berlin-Frankfurt-Münster (iBFM)-FLOW-networkthe AutoFLOW consortium. Automated Flow Cytometric MRD Assessment in Childhood Acute B- Lymphoblastic Leukemia Using Supervised Machine Learning. Cytometry A 2019 Sep 07;95(9):966-975 [FREE Full text] [doi: 10.1002/cyto.a.23852] [Medline: 31282025]
75. Gal O, Auslander N, Fan Y, Meerzaman D. Predicting Complete Remission of Acute Myeloid Leukemia: Machine Learning Applied to Gene Expression. Cancer Inform 2019;18:1176935119835544 [FREE Full text] [doi: 10.1177/1176935119835544] [Medline: 30911218]
76. Castillo D, Galvez JM, Herrera LJ, Rojas F, Valenzuela O, Caba O, et al. Leukemia multiclass assessment and classification from Microarray and RNA-seq technologies integration at gene expression level. PLoS One 2019;14(2):e0212127 [FREE Full text] [doi: 10.1371/journal.pone.0212127] [Medline: 30753220]
77. Li H, Xu C, Xin B, Zheng C, Zhao Y, Hao K, et al. F-FDG PET/CT Radiomic Analysis with Machine Learning for Identifying Bone Marrow Involvement in the Patients with Suspected Relapsed Acute Leukemia. Theranostics 2019;9(16):4730-4739 [FREE Full text] [doi: 10.7150/thno.33841] [Medline: 31367253]
78. Parida L, Haferlach C, Rhrissorrakrai K, Utro F, Levovitz C, Kern W, et al. Dark-matter matters: Discriminating subtle blood cancers using the darkest DNA. PLoS Comput Biol 2019 Aug;15(8):e1007332 [FREE Full text] [doi: 10.1371/journal.pcbi.1007332] [Medline: 31469830]
79. Mallesh N, Zhao M, Elsner F, Lueling H, Schabath R, Haferlach C. Knowledge transfer between artificial neural networks for different multicolor flow cytometry protocols improves classification performance for rare B-cell neoplasm subtypes. Blood 2019;134(Supplement_1):2094. [doi: 10.1182/blood-2019-130483]
80. Pansombut T, Wikaisuksakul S, Khongkraphan K, Phon-On A. Convolutional Neural Networks for Recognition of Lymphoblast Cell Images. Comput Intell Neurosci 2019;2019:7519603 [FREE Full text] [doi: 10.1155/2019/7519603] [Medline: 31281337]
81. Mosquera Orgueira A, Antelo Rodríguez B, Alonso Vence N, Bendaña López Á, Díaz Arias JÁ, Díaz Varela N, et al. Time to Treatment Prediction in Chronic Lymphocytic Leukemia Based on New Transcriptional Patterns. Front Oncol 2019 Feb 15;9:79 [FREE Full text] [doi: 10.3389/fonc.2019.00079] [Medline: 30828568]
82. Al-Fahad R, Yeasin M, Glass J, Conklin H, Jacola L, Reddick W. Early Imaging-Based Predictive Modeling of Cognitive Performance Following Therapy for Childhood ALL. IEEE Access 2019;7:146662-146674. [doi: 10.1109/ACCESS.2019.2946240]
83. Syed-Abdul S, Firdani RP, Chung HJ, Uddin M, Hur M, Park JH, et al. Author Correction: Artificial Intelligence based Models for Screening of Hematologic Malignancies using Cell Population Data. Sci Rep 2020 Jun 23;10(1):10475 [FREE Full text] [doi: 10.1038/s41598-020-67559-5] [Medline: 32572136]
84. Deeb SJ, Tyanova S, Hummel M, Schmidt-Supprian M, Cox J, Mann M. Machine Learning-based Classification of Diffuse Large B-cell Lymphoma Patients by Their Protein Expression Profiles. Mol Cell Proteomics 2015 Nov;14(11):2947-2960 [FREE Full text] [doi: 10.1074/mcp.M115.050245] [Medline: 26311899]
85. Negm AS, Hassan OA, Kandil AH. A decision support system for Acute Leukaemia classification based on digital microscopic images. Alexandria Engineering Journal 2018 Dec;57(4):2319-2332. [doi: 10.1016/j.aej.2017.08.025]
86. Moshavash Z, Danyali H, Helfroush MS. An Automatic and Robust Decision Support System for Accurate Acute Leukemia Diagnosis from Blood Microscopic Images. J Digit Imaging 2018 Oct;31(5):702-717 [FREE Full text] [doi: 10.1007/s10278-018-0074-y] [Medline: 29654425]

87. Muhsen IN, Shyr D, Sung AD, Hashmi SK. Machine Learning Applications in the Diagnosis of Benign and Malignant Hematological Diseases. Clin Hematol Int 2021 Mar;3(1):13-20 [FREE Full text] [doi: 10.2991/chi.k.201130.001] [Medline: 34595462]
88. Agarwal N, Agarwal P. Early Stage Detection of Leukemia Using Artificial Intelligence. In: Mohanty SN, Nalinipriya G, Jena OP, Sarkar A, editors. Machine Learning for Healthcare Applications. Hoboken, NJ: Wiley; Apr 12, 2021:215-224.
89. Bano S, Rao K. Partial Context Similarity of Gene/Proteins in Leukemia Using Context Rank Based Hierarchical Clustering Algorithm. IJECE 2015 Jun 01;5(3):483. [doi: 10.11591/ijece.v5i3.pp483-490]
90. Ng D, Wu D, Wood B, Fromm J. Computer-aided detection of rare tumor populations in flow cytometry: an example with classic Hodgkin lymphoma. Am J Clin Pathol 2015 Sep;144(3):517-524. [doi: 10.1309/AJCPY8E2LYHCGUFP] [Medline: 26276783]
91. Sirinukunwattana K, Aberdeen A, Theissen H, Sousos N, Psaila B, Mead A, et al. Artificial intelligence-based morphological fingerprinting of megakaryocytes: a new tool for assessing disease in MPN patients. Blood Adv 2020 Jul 28;4(14):3284-3294 [FREE Full text] [doi: 10.1182/bloodadvances.2020002230] [Medline: 32706893]
92. Carreras J, Kikuti Y, Miyaoka M, Hiraiwa S, Tomita S, Ikoma H, et al. A Single Gene Expression Set Derived from Artificial Intelligence Predicted the Prognosis of Several Lymphoma Subtypes; and High Immunohistochemical Expression of TNFAIP8 Associated with Poor Prognosis in Diffuse Large B-Cell Lymphoma. AI 2020 Jul 21;1(3):342-360. [doi: 10.3390/ai1030023]
93. Arai Y, Kondo T, Fuse K, Shibasaki Y, Masuko M, Sugita J. Prediction of acute graft-versus-host disease following allogeneic hematopoietic stem cell transplantation using a machine learning algorithm. Blood 2018;132(Supplement 1):68. [doi: 10.1182/blood-2018-99-114794]
94. Hamey FK, Göttgens B. Machine learning predicts putative hematopoietic stem cells within large single-cell transcriptomics data sets. Exp Hematol 2019 Oct;78:11-20 [FREE Full text] [doi: 10.1016/j.exphem.2019.08.009] [Medline: 31513832]
95. Ko B, Wang Y, Li J, Li C, Weng P, Hsu S, et al. Clinically validated machine learning algorithm for detecting residual diseases with multicolor flow cytometry analysis in acute myeloid leukemia and myelodysplastic syndrome. EBioMedicine 2018 Nov;37:91-100 [FREE Full text] [doi: 10.1016/j.ebiom.2018.10.042] [Medline: 30361063]
96. Chen D, Goyal G, Go RS, Parikh SA, Ngufor CG. Improved Interpretability of Machine Learning Model Using Unsupervised Clustering: Predicting Time to First Treatment in Chronic Lymphocytic Leukemia. JCO Clin Cancer Inform 2019 May;3:1-11 [FREE Full text] [doi: 10.1200/CCI.18.00137] [Medline: 31112417]
97. Gunčar G, Kukar M, Notar M, Brvar M, Černelč P, Notar M, et al. An application of machine learning to haematological diagnosis. Sci Rep 2018 Jan 11;8(1):411 [FREE Full text] [doi: 10.1038/s41598-017-18564-8] [Medline: 29323142]
98. Elemento O. Toward artificial intelligence driven pathology assessment for hematologic malignancies. Blood Cancer Discovery 2021;2(3):195-197. [doi: 10.1158/2643-3230.bcd-21-0048]
99. Kaestner L. Artificial intelligence meets hematology. Transfus Apher Sci 2020 Dec;59(6):102986. [doi: 10.1016/j.transci.2020.102986] [Medline: 33153901]
100. Shreve J, Meggendorfer M, Awada H, Mukherjee S, Walter W, Hutter S. A personalized prediction model to risk stratify patients with acute myeloid leukemia (AML) using Artificial Intelligence. Blood 2019;134(Supplement_1):2091. [doi: 10.1182/blood-2019-128066]
101. Siddiqui N, Klein A, Godara A, Varga C, Buchsbaum R, Hughes M. Supervised machine learning algorithms using patient related factors to predict in-hospital mortality following acute myeloid leukemia therapy. Blood 2019;134(Supplement_1):3435. [doi: 10.1182/blood-2019-128823]
102. Shouval R, Labopin M, Gorin NC, Bomze D, Houhou M, Blaise D, et al. Individualized prediction of leukemia-free survival after autologous stem cell transplantation in acute myeloid leukemia. Cancer 2019 Oct 15;125(20):3566-3573 [FREE Full text] [doi: 10.1002/cncr.32344] [Medline: 31225904]
103. Arai Y, Kondo T, Fuse K, Shibasaki Y, Masuko M, Sugita J, et al. Using a machine learning algorithm to predict acute graft-versus-host disease following allogeneic transplantation. Blood Adv 2019 Nov 26;3(22):3626-3634 [FREE Full text] [doi: 10.1182/bloodadvances.2019000934] [Medline: 31751471]
104. Morita K, Wang F, Makishima H, Yan Y, Yoshizato T, Yoshida K. Pan-Myeloid Leukemia Analysis: Machine learning-based approach to predict phenotype and clinical outcomes using mutation data. Blood 2018;132(Supplement 1):1801. [doi: 10.1182/blood-2018-99-116685]
105. Shouval R, Bonifazi F, Fein J, Boschini C, Oldani E, Labopin M, et al. Validation of the acute leukemia-EBMT score for prediction of mortality following allogeneic stem cell transplantation in a multi-center GITMO cohort. Am J Hematol 2017 May;92(5):429-434 [FREE Full text] [doi: 10.1002/ajh.24677] [Medline: 28181279]
106. Ananda-Rajah MR, Bergmeir C, Petitjean F, Slavin MA, Thursky KA, Webb GI. Toward Electronic Surveillance of Invasive Mold Diseases in Hematology-Oncology Patients: An Expert System Combining Natural Language Processing of Chest Computed Tomography Reports, Microbiology, and Antifungal Drug Data. JCO Clin Cancer Inform 2017 Nov;1:1-10 [FREE Full text] [doi: 10.1200/CCI.17.00011] [Medline: 30657390]
107. Fleming S, Tsai C, Döhner H, Döhner K, Papaemmanuil E. Use of Machine Learning in 2074 Cases of Acute Myeloid Leukemia for Genetic Risk Profiling. Blood 2019;134(Supplement_1):1392. [doi: 10.1182/blood-2019-128243]

108. Sheikhpour R, Fazli R, Mehrabani S. Gene Identification from Microarray Data for Diagnosis of Acute Myeloid and Lymphoblastic Leukemia Using a Sparse Gene Selection Method. IJPHO 2021 Mar 28;11(2):70-77. [doi: 10.18502/ijpho.v11i2.5838]
109. Patmore R, Smith A, Cathy B, Roman E, Crouch S. Predicting time to treatment in follicular lymphoma using population-based data. Blood 2015;126(23):3912. [doi: 10.1182/blood.v126.23.3912.3912]
110. Rezayi S, Mohammadzadeh N, Bouraghi H, Saeedi S, Mohammadpour A. Timely Diagnosis of Acute Lymphoblastic Leukemia Using Artificial Intelligence-Oriented Deep Learning Methods. Comput Intell Neurosci 2021;2021:5478157 [FREE Full text] [doi: 10.1155/2021/5478157] [Medline: 34804144]
111. Nanaa A, Akkus Z, Lee WY, Pantanowitz L, Salama ME. Machine learning and augmented human intelligence use in histomorphology for haematolymphoid disorders. Pathology 2021 Apr;53(3):400-407. [doi: 10.1016/j.pathol.2020.12.004] [Medline: 33642096]
112. Boldú L, Merino A, Alférez S, Molina A, Acevedo A, Rodellar J. Automatic recognition of different types of acute leukaemia in peripheral blood by image analysis. J Clin Pathol 2019 Nov 29;72(11):755-761. [doi: 10.1136/jclinpath-2019-205949] [Medline: 31256009]
113. Ng D, Zuromski L. Augmented Human Intelligence and Automated Diagnosis in Flow Cytometry for Hematologic Malignancies. Am J Clin Pathol 2021 Mar 15;155(4):597-605. [doi: 10.1093/ajcp/aqaa166] [Medline: 33210119]
114. Sasaki K, Jabbour EJ, Ravandi F, Konopleva M, Borthakur G, Wierda WG, et al. The LEukemia Artificial Intelligence Program (LEAP) in chronic myeloid leukemia in chronic phase: A model to improve patient outcomes. Am J Hematol 2021 Feb 01;96(2):241-250 [FREE Full text] [doi: 10.1002/ajh.26047] [Medline: 33180322]
115. Fauzi IR, Rustam Z, Wibowo A. Multiclass classification of leukemia cancer data using Fuzzy Support Vector Machine (FSVM) with feature selection using Principal Component Analysis (PCA). J. Phys.: Conf. Ser 2021 Jan 01;1725(1):012012 [FREE Full text] [doi: 10.1088/1742-6596/1725/1/012012]
116. Nikitaev VG, Tupitsyn NN, Pronichev AN, Polyakov EV, Dmitrieva VV, Chernysheva OA, et al. Artificial Intelligence Technologies in the Diagnosis of Acute Lymphoblastic Leukemia and Minimal Residual Disease. Biomed Eng 2021 Jan 06;54(5):354-356. [doi: 10.1007/s10527-021-10038-6]
117. Mou A, Hasan M, Saha P, Priom N, Saha A. Prediction and Rule Generation for Leukemia using Decision Tree and Association Rule Mining. 2021 Apr 05 Presented at: 2020 11th International Conference on Electrical and Computer Engineering (ICECE); December 17-19, 2020; Dhaka, Bangladesh p. 133-136. [doi: 10.1109/icece51571.2020.9393069]
118. Li C, Zhu B, Chen J, Huang X. Feature genes predicting the FLT3/ITD mutation in acute myeloid leukemia. Mol Med Rep 2016 Jul;14(1):89-94 [FREE Full text] [doi: 10.3892/mmr.2016.5260] [Medline: 27177049]
119. Ayyappan V, Chang A, Zhang C, Paidi SK, Bordett R, Liang T, et al. Identification and Staging of B-Cell Acute Lymphoblastic Leukemia Using Quantitative Phase Imaging and Machine Learning. ACS Sens 2020 Oct 23;5(10):3281-3289. [doi: 10.1021/acssensors.0c01811] [Medline: 33092347]
120. Fathi E, Rezaee MJ, Tavakkoli-Moghaddam R, Alizadeh A, Montazer A. Design of an integrated model for diagnosis and classification of pediatric acute leukemia using machine learning. Proc Inst Mech Eng H 2020 Oct;234(10):1051-1069. [doi: 10.1177/0954411920938567] [Medline: 32633668]
121. Al-Tahhan FE, Fares ME, Sakr AA, Aladle DA. Accurate automatic detection of acute lymphatic leukemia using a refined simple classification. Microsc Res Tech 2020 Oct;83(10):1178-1189. [doi: 10.1002/jemt.23509] [Medline: 32497337]
122. Miyoshi H, Sato K, Kabeya Y, Yonezawa S, Nakano H, Takeuchi Y, et al. Deep learning shows the capability of high-level computer-aided diagnosis in malignant lymphoma. Lab Invest 2020 Oct;100(10):1300-1310. [doi: 10.1038/s41374-020-0442-3] [Medline: 32472096]
123. Nazari E, Farzin AH, Aghemiri M, Avan A, Tara M, Tabesh H. Deep Learning for Acute Myeloid Leukemia Diagnosis. J Med Life 2020;13(3):382-387 [FREE Full text] [Medline: 33072212]
124. Özyurt F. A fused CNN model for WBC detection with MRMR feature selection and extreme learning machine. Soft Comput 2019 Sep 26;24(11):8163-8172. [doi: 10.1007/s00500-019-04383-8]
125. Shafique S, Tehsin S. Acute Lymphoblastic Leukemia Detection and Classification of Its Subtypes Using Pretrained Deep Convolutional Neural Networks. Technol Cancer Res Treat 2018 Jan 01;17:1533033818802789 [FREE Full text] [doi: 10.1177/1533033818802789] [Medline: 30261827]
126. Madhu G. Computer Vision and machine learning approach for malaria diagnosis in thin blood smears from microscopic blood images. In: Machine Learning for Intelligent Decision Science. Singapore City, Singapore: Springer; Apr 03, 2020:191-209.
127. Li D, Bledsoe JR, Zeng Y, Liu W, Hu Y, Bi K, et al. A deep learning diagnostic platform for diffuse large B-cell lymphoma with high accuracy across multiple hospitals. Nat Commun 2020 Nov 26;11(1):6004 [FREE Full text] [doi: 10.1038/s41467-020-19817-3] [Medline: 33244018]
128. Sheikh I, Chachoo M. Advanced machine learning for leukaemia detection based on White Blood Cell Segmentation. 2020 Nov 05 Presented at: International Conference on Information, Communication and Computing Technology; May 8, 2020; New Delhi, India p. 195-207. [doi: 10.1007/978-981-15-9671-1_17]
129. Mohammed EA, Mohamed MM, Naugler C, Far BH. Toward leveraging big value from data: chronic lymphocytic leukemia cell classification. Netw Model Anal Health Inform Bioinforma 2017 Feb 18;6(1):e1. [doi: 10.1007/s13721-017-0146-9]

130. Su J, Liu S, Song J. A segmentation method based on HMRF for the aided diagnosis of acute myeloid leukemia. Comput Methods Programs Biomed 2017 Dec;152:115-123. [doi: 10.1016/j.cmpb.2017.09.011] [Medline: 29054251]
131. Pan L, Liu G, Lin F, Zhong S, Xia H, Sun X, et al. Machine learning applications for prediction of relapse in childhood acute lymphoblastic leukemia. Sci Rep 2017 Aug 07;7(1):7402 [FREE Full text] [doi: 10.1038/s41598-017-07408-0] [Medline: 28784991]
132. Mohajerani P, Ntziachristos V. Classification of Normal Versus Malignant Cells in B-ALL Microscopic Images Based on a Tiled Convolution Neural Network Approach. In: ISBI 2019 C-NMC Challenge: Classification in Cancer Cell Imaging. Singapore City, Singapore: Springer; 2019:103-111.
133. Moraes LO, Pedreira CE, Barrena S, Lopez A, Orfao A. A decision-tree approach for the differential diagnosis of chronic lymphoid leukemias and peripheral B-cell lymphomas. Comput Methods Programs Biomed 2019 Sep;178:85-90 [FREE Full text] [doi: 10.1016/j.cmpb.2019.06.014] [Medline: 31416565]
134. Boldú L, Merino A, Acevedo A, Molina A, Rodellar J. A deep learning model (ALNet) for the diagnosis of acute leukaemia lineage using peripheral blood cell images. Comput Methods Programs Biomed 2021 Apr;202:105999. [doi: 10.1016/j.cmpb.2021.105999] [Medline: 33618145]
135. Coombes C, Abrams Z, Li S, Abruzzo L, Coombes K. Unsupervised machine learning and prognostic factors of survival in chronic lymphocytic leukemia. J Am Med Inform Assoc 2020 Jul 01;27(7):1019-1027 [FREE Full text] [doi: 10.1093/jamia/ocaa060] [Medline: 32483590]
136. Warnat-Herresthal S, Perrakis K, Taschler B, Becker M, Baßler K, Beyer M, et al. Scalable Prediction of Acute Myeloid Leukemia Using High-Dimensional Machine Learning and Blood Transcriptomics. iScience 2020 Jan 24;23(1):100780 [FREE Full text] [doi: 10.1016/j.isci.2019.100780] [Medline: 31918046]
137. Bachar N, Benbassat D, Brailovsky D, Eshel Y, Glück D, Levner D, et al. An artificial intelligence-assisted diagnostic platform for rapid near-patient hematology. Am J Hematol 2021 Oct 01;96(10):1264-1274. [doi: 10.1002/ajh.26295] [Medline: 34264525]
138. Bornhauser B, Boutter J, Horvath P, Stanulla M. Image-based RNA interference screening identifies Microenvironmental signals supporting primary acute lymphoblastic leukemia cell survival. Blood 2012;120(21):2348. [doi: 10.1182/blood.v120.21.2348.2348]
139. Harkins RA, Chang A, Patel SP, Lee MJ, Goldstein JS, Merdan S, et al. Remaining challenges in predicting patient outcomes for diffuse large B-cell lymphoma. Expert Rev Hematol 2019 Nov;12(11):959-973 [FREE Full text] [doi: 10.1080/17474086.2019.1660159] [Medline: 31513757]
140. Sarno J, Davis KL. Single-cell mass cytometry and machine learning predict relapse in childhood leukemia. Mol Cell Oncol 2018;5(5):e1472057 [FREE Full text] [doi: 10.1080/23723556.2018.1472057] [Medline: 30263942]
141. Tyler E, Jungbluth A, O'Reilly R, Koehne G. WT1-specific T-cell responses in high-risk multiple myeloma patients undergoing allogeneic T cell-depleted hematopoietic stem cell transplantation and donor lymphocyte infusions. Blood 2013 Jan 10;121(2):308-317 [FREE Full text] [doi: 10.1182/blood-2012-06-435040] [Medline: 23160468]
142. Sivapalaratnam S. Artificial intelligence and machine learning in haematology. Br J Haematol 2019 Apr;185(2):207-208. [doi: 10.1111/bjh.15774] [Medline: 30729496]
143. Mahmood N, Shahid S, Bakhshi T, Riaz S, Ghufran H, Yaqoob M. Identification of significant risks in pediatric acute lymphoblastic leukemia (ALL) through machine learning (ML) approach. Med Biol Eng Comput 2020 Nov;58(11):2631-2640. [doi: 10.1007/s11517-020-02245-2] [Medline: 32840766]
144. Hauser R, Esserman D, Beste L, Ong S, Colomb D, Bhargava A, et al. A Machine Learning Model to Successfully Predict Future Diagnosis of Chronic Myelogenous Leukemia With Retrospective Electronic Health Records Data. Am J Clin Pathol 2021 Nov 08;156(6):1142-1148. [doi: 10.1093/ajcp/aqab086] [Medline: 34184028]
145. Steyerberg EW. Clinical Prediction Models A Practical Approach to Development, Validation, and Updating. In: Statistics for Biology and Health. Berlin, Germany: Springer; 2020.

Abbreviations

AI: artificial intelligence
ALL: acute lymphoblastic leukemia
AML: acute myeloid leukemia
ANN: artificial neural network
AUC: area under the curve
CBC: complete blood count
CLL: chronic lymphocytic leukemia
CLL-TIM: CLL Treatment-Infection Model
CML: chronic myeloid leukemia
CNN: convolutional neural network
CR: complete remission
DL: deep learning

DNN: deep neural network
GOTDP-MP-CNNs: globally optimized transfer deep-learning platform with multiple pretrained CNNs
HSCRKM: histogram-based soft covering rough K-means clustering
IPI: International Prognostic Index
ML: machine learning
RBC: red blood cell
RF: random forest
ROC: receiver operating characteristic
SEER: Surveillance, Epidemiology, and End Results
SESSA: statistically enhanced Salp Swarm Algorithm
SOM: self-organizing map
SVM: support vector machine
WBC: white blood cell

Edited by R Kukafka, G Eysenbach; submitted 16.01.22; peer-reviewed by JA Benítez-Andrades, B Puladi, K Rahmani; comments to author 19.03.22; revised version received 14.05.22; accepted 29.05.22; published 12.07.22.

Please cite as:
El Alaoui Y, Elomri A, Qaraqe M, Padmanabhan R, Yasin Taha R, El Omri H, EL Omri A, Aboumarzouk O
A Review of Artificial Intelligence Applications in Hematology Management: Current Practices and Future Prospects
J Med Internet Res 2022;24(7):e36490
URL: https://www.jmir.org/2022/7/e36490
doi:10.2196/36490
PMID:35819826

Review

Digital Facilitation to Support Patient Access to Web-Based Primary Care Services: Scoping Literature Review

Brandi Leach[1], PhD; Sarah Parkinson[1], MSc; Evangelos Gkousis[1], MSc; Gary Abel[2], PhD; Helen Atherton[3], MPH, PhD; John Campbell[2], MBE, MD; Christopher Clark[2], PhD; Emma Cockcroft[2], PhD; Christine Marriott[4]; Emma Pitchforth[2], PhD; Jon Sussex[1], MSc

[1]RAND Europe, Cambridge, United Kingdom

[2]University of Exeter Medical School, Exeter, United Kingdom

[3]University of Warwick, Coventry, United Kingdom

[4]National Institute of Health and Care Research Collaboration South West Peninsula Patient Engagement Group, University of Exeter Medical School, Exeter, United Kingdom

Corresponding Author:
Brandi Leach, PhD
RAND Europe
Westbrook Centre
Cambridge, CB41YG
United Kingdom
Phone: 44 1223353329
Email: bleach@randeurope.org

Abstract

Background: The use of web-based services within primary care (PC) in the National Health Service in England is increasing, with medically underserved populations being less likely to engage with web-based services than other patient groups. *Digital facilitation*—referring to a range of processes, procedures, and personnel that seek to support patients in the uptake and use of web-based services—may be a way of addressing these challenges. However, the models and impact of digital facilitation currently in use are unclear.

Objective: This study aimed to identify, characterize, and differentiate between different approaches to digital facilitation in PC; establish what is known about the effectiveness of different approaches; and understand the enablers of digital facilitation.

Methods: Adopting scoping review methodology, we searched academic databases (PubMed, EMBASE, CINAHL, Web of Science, and Cochrane Library) and gray literature published between 2015 and 2020. We conducted snowball searches of reference lists of included articles and articles identified during screening as relevant to digital facilitation, but which did not meet the inclusion criteria because of article type restrictions. Titles and abstracts were independently screened by 2 reviewers. Data from eligible studies were analyzed using a narrative synthesis approach.

Results: A total of 85 publications were included. Most (71/85, 84%) were concerned with digital facilitation approaches targeted at patients (promotion of services, training patients to improve their technical skills, or other guidance and support). Further identified approaches targeted PC staff to help patients (eg, improving staff knowledge of web-based services and enhancing their technical or communication skills). Qualitative evidence suggests that some digital facilitation may be effective in promoting the uptake and use of web-based services by patients (eg, recommendation of web-based services by practice staff and coaching). We found little evidence that providing patients with initial assistance in registering for or accessing web-based services leads to increased long-term use. Few studies have addressed the effects of digital facilitation on health care inequalities. Those that addressed this suggested that providing technical training for patients could be effective, at least in part, in reducing inequalities, although not entirely. Factors affecting the success of digital facilitation include perceptions of the usefulness of the web-based service, trust in the service, patients' trust in providers, the capacity of PC staff, guidelines or regulations supporting facilitation efforts, and staff buy-in and motivation.

Conclusions: Digital facilitation has the potential to increase the uptake and use of web-based services by PC patients. Understanding the approaches that are most effective and cost-effective, for whom, and under what circumstances requires further research, including rigorous evaluations of longer-term impacts. As efforts continue to increase the use of web-based services in PC in England and elsewhere, we offer an early typology to inform conceptual development and evaluations.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42020189019; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=189019

(J Med Internet Res 2022;24(7):e33911) doi:10.2196/33911

KEYWORDS

web-based health services; primary care; digital facilitation

Introduction

Background

The use of web-based services within primary care in the National Health Service (NHS) in England is increasing, with 33% of patients registered to use at least one web-based service in January 2021 compared with only 19% in April 2017 [1]. Although still at levels below those found in other countries such as the United States [2], the use of web-based services is likely to grow, given that it is supported by contractual mandates from NHS England [3]; it is part of wider efforts to establish digitally enabled care [4]; and because of increased pressure on health care services, technological progress, and changing public expectations [5]. The use of web-based services has accelerated across primary care in many countries during the COVID-19 pandemic as a means of enabling distanced care [6-10]. Within the NHS, services provided by all primary care practices include booking a consultation (via a practice website or through a web platform linked to a practice website), ordering repeat prescriptions, and accessing electronic health records. Additional services include secure messaging, provision of test results, having a consultation (ie, receiving a response from the practice via SMS text messages, web-based messages, phone calls, or video calls), facilitating access to external resources (eg, referring patients to websites or apps that can augment their care), and providing access to practice websites for informational purposes.

The increased use of web-based services has been shown to benefit patients, general medical practitioners (primary care physicians, known as *general practitioners [GPs]* in the United Kingdom), and other primary care staff through improved communication between patients and GP practices, expanded health and health care knowledge for patients, and improved access to services [11-13]. For GP practices and patients to gain the potential benefits that technological innovation can bring to primary care, patients must be able to, as well as wish to, access and use web-based services [14]. There is emerging evidence that the trend toward web-based interactions creates or exacerbates pre-existing inequalities in access to health care information and services for some patient groups who may not be able, or may not choose, to use or access web-based services [15,16].

A way of supporting the use of web-based services and countering the potential for increasing inequalities may be through *digital facilitation*, which we have defined as "that range of processes, procedures, and personnel which seeks to support NHS patients in their uptake and use of online services" [17]. Digital facilitation ranges from the promotion of web-based services on a practice website to active coaching in the use of web-based services and provision of training and education to practice staff in the use of services so that they can better assist patients [18]. For the purposes of this research, we have not extended the scope of *digital facilitation* to include the facilitation of access to digitally based therapeutic interventions.

Medically underserved and vulnerable populations are less likely than other patient groups to engage in web-based services [2,19]. The reasons for lower engagement in web-based services among medically underserved populations are complex. They include factors focusing on limited access to services (eg, poor internet connection), as well as those affecting motivations to engage (eg, lack of familiarity with the internet, lower health or computer literacy, and lack of trust in web-based information sources) [19-22]. It has been suggested that a way of reducing inequalities related to the use of web-based services in health care may be to actively support vulnerable population groups in accessing and using web-based services through digital facilitation [23]. The digital competence of health care professionals and their acceptance of web-based service provision are also important for the successful implementation of web-based patient services.

Objectives

Recognizing the lack of understanding of digital facilitation and its role in supporting the use of web-based services in primary care, we conducted a systematic scoping review. We aimed to identify, characterize, and differentiate between different approaches to digital facilitation in primary care; create a typology of these approaches; establish what is known about the effectiveness, perceived advantages, and challenges of different approaches to digital facilitation; examine how they affect inequalities of access to web-based services; and explore factors enabling digital facilitation. We also sought early indications of the extent to which the COVID-19 pandemic might be associated with changing approaches to digital facilitation.

Methods

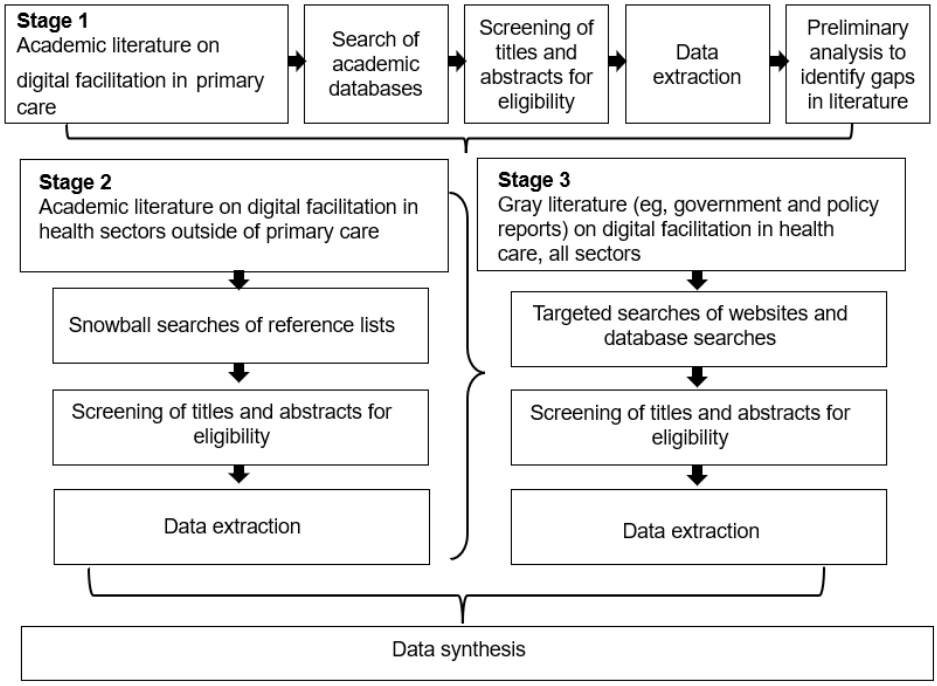
Overview

We conducted a systematic scoping review of the literature. Scoping reviews are appropriate for clarifying conceptual boundaries on topics, such as digital facilitation, where a concept is new and poorly defined in the literature [24]. The scoping review was conducted in stages (Figure 1) to allow learning from earlier stages to be fed into later stages. The protocol for the study was registered with PROSPERO (International Prospective Register of Systematic Reviews; registration number CRD42020189019) [25].

Our focus is on digital facilitation within primary care in England; however, we also consider digital facilitation in other

geographical areas and other health care sectors where there is clear relevance to primary care. Primary care is distinct from other types of health care in that it is typically the patient's first point of contact within the health system, and the primary care staff is tasked with caring for the patient as a whole rather than focusing on specific conditions [26]. Primary care is also at the center of the NHS's Digital First plans [27] and faces particular challenges around rising demands in the face of workforce pressures [28]. Although this study focuses on primary care, some findings will also be relevant to wider health care contexts.

Figure 1. Flow of the literature review process.



Patient and Public Involvement

This review was conducted in collaboration with a study-specific patient advisory group. The group included patients and caregivers. A total of 2 web-based meetings took place over the course of the project; between meetings, group members were involved via email or on a one-on-one basis. The patient advisory group contributed to the development of the search strategy; operationalization of key terms; discussion of findings, including identified themes and gaps; data synthesis; and report drafting.

Searches

Stage 1: Academic Literature on Digital Facilitation in Primary Care

We searched the following databases: PubMed, EMBASE, CINAHL, Web of Science, and Cochrane Library. The search strategy focused on three key concepts: (1) web-based services, (2) digital facilitation, and (3) primary care settings. We restricted the searches to the European Economic Area and Organization for Economic Cooperation and Development countries as these would likely be most relevant to primary care practices in England. Full details of the search strategy are available in Multimedia Appendix 1. The first round of stage 1 searches covered publications from January 2010 to June 2020; however, these were restricted during pilot screening from January 2015 to June 2020 (see the *Study Selection* section).

Stage 2: Snowball Searches to Identify Literature on Digital Facilitation in Health Sectors Outside of Primary Care

In stage 2, we screened the reference lists of all articles identified for inclusion in stage 1, in addition to the reference lists of articles that we identified during stage 1 as not fitting the inclusion criteria because of article type restrictions (eg, protocols or editorials) but which were otherwise relevant.

Stage 3: Gray Literature on Digital Facilitation in Health Care

Gray literature was searched to identify relevant government and policy institute reports on digital facilitation in health care. This involved searches of 3 relevant not-for-profit research institutes (The Health Foundation, The King's Fund, and The Nuffield Trust) and a health professional association website (Royal College of General Practitioners), as well as a general search of the Health Management Information Consortium database. The targeted searches of websites used combinations of terms such as *online services*, *digital*, *access*, and *patients* using Boolean operators where website search functions allowed it. The Health Management Information Consortium database allowed more complex searches; therefore, we adopted a search strategy that captured concepts related to the web (eg, web-based, digital, internet-based, and technology) and facilitation (eg, uptake, encouragement, and increased use). Full details of the gray literature search strategy are available in Multimedia Appendix 1.

Additional Searches Not Included in Final Review

We also explored some academic literature on digital facilitation in non–health care sectors similar to primary care in that they incorporated both web-based and offline customer services (ie, tourism and travel and retail banking) to see if any methods of digital facilitation were mentioned there that were not covered in the health care literature. These searches did not reveal additional approaches to digital facilitation and are not reported here.

Study Selection

A key inclusion criterion for all the publications was that they addressed the facilitation of web-based services. Defining the inclusion criteria a priori was challenging, given that the key aim of this work was to define the scope of digital facilitation. We focused on web-based services that were accessed by patients through websites or phone apps facilitating access to care or providing resources for self-care and not on the delivery of medical therapies through web-based platforms, such as web-based mental health therapy. These services reflected those supported by primary care practices and were in line with the focus of NHS England at the time of this research [3]. We operationalized digital facilitation and web-based services as detailed in Textbox 1. Further eligibility criteria were tailored to the stage of the screening process (eg, primary care literature and nonprimary health care literature). Detailed inclusion and exclusion criteria for each stage of the screening process are presented in Table 1.

Before the full screening of the 11,853 publications from stage 1, we undertook a pilot screening exercise examining 237 (2%) publications, during which publications were jointly screened by 2 reviewers (EG and SP) and the results were discussed to ensure consistent approaches to screening. During the pilot screening, it was agreed that publications for stages 1 and 3 would be restricted to articles published from 2015 to 2020. For stage 2, which relied on snowball-type searches of reference lists, we included articles from 2010 to 2020 as the reference lists would have had few or no eligible articles if we did not expand the inclusion criteria to earlier years. Following the pilot screening, all remaining publications were screened independently by 1 of 2 reviewers (EG and SP).

Textbox 1. Operationalization of digital facilitation and web-based services.

Concept and inclusion and exclusion criteria

Inclusion criteria

- Digital facilitation: papers that included reference to what was done to help patients access and use web-based services, including (but not limited to) the following:
  - In-person assistance with using web-based services
  - Active methods of web-based assistance for accessing services (eg, chat or help functions)
  - Passive methods of web-based assistance for accessing services (eg, frequently asked questions and help pages)
  - Telephone-based methods of providing assistance for accessing services (eg, helplines)
  - Public awareness campaigns around web-based services (if done by general practices)
  - Service improvements if done explicitly to improve or increase access
- Web-based services: web-based services accessed through a website or app, such as the following:
  - Health records
  - Prescription ordering
  - Appointment booking
  - eConsult or other web-based methods used to triage patients
  - Health care information

Exclusion criteria

- Digital facilitation: Papers without information on what was done to help patients access and use web-based services
- Web-based services: Non–web-based services (eg, telephone only), wearable devices, delivery of therapies (eg, psychotherapies) on the web, and web-based services for general practitioners or physicians, which did not include patients (eg, accessing continuing medical education and web-based clinical decision support tools without input from the patient)

Table 1. Inclusion and exclusion criteria for the screening process.

Stage of process and criteria	Include	Exclude
All stages		
Scale and spread of intervention	At all scales and geographic levels from the individual site to national coverage	None
Country	EEA[a] countries or non-European high-income countries (defined as membership in OECD[b])	Countries not in the EEA or OECD
Language	English	Languages other than English
Availability	Full-text availability	Title and abstract only and conference proceedings with no full-text article
Stages 1 and 3		
Year of publication	2015 to January 2020	2014 or earlier
Stage 2		
Year of publication	2010 to January 2020	2009 or earlier
Stage 1 only: screening of academic literature on digital facilitation in primary care		
Topic relevance	Digital facilitation of web-based services in primary health care settings; where digital facilitation was implemented in some form: implementation as part of routine service delivery or implementation for research purposes	Where there was no reference to facilitation being implemented by or on behalf of primary care practices; thus, solely theoretical papers were excluded
Article type	Original research	Theoretical and commentary articles; trial registrations (ie, articles registered on ClinicalTrials.gov or the WHO ICTRP[c] registry)
Stage 2 only: screening of literature on digital facilitation in health sectors outside of primary care		
Topic relevance	Digital facilitation of web-based services in non–primary care health sectors; where digital facilitation was implemented in some form: implementation as part of routine service delivery or implementation for research purposes; articles addressing aspects of digital facilitation found not to be covered by articles identified in stage 1; key gaps include evaluations of digital facilitation approaches, cost-effectiveness, and effectiveness of digital facilitation approaches for vulnerable populations	Where there was no reference to facilitation being implemented by or on behalf of health care providers; thus, solely theoretical papers were excluded; articles addressing aspects of digital facilitation already covered by the included articles identified in stage 1
Article type	Original research	Theoretical and commentary articles and trial registrations (ie, articles registered on ClinicalTrials.gov or the WHO ICTRP registry)
Stage 3 only: screening of gray literature on digital facilitation in health care, all sectors		
Topic relevance	Digital facilitation of web-based services in health care, all sectors; articles addressing aspects of digital facilitation found not to be covered by articles identified in stage 1; key gaps include the following: implications of COVID-19 pandemic for digital facilitation, evaluations of digital facilitation approaches, and effectiveness of digital facilitation approaches for vulnerable populations	Where there was no reference to facilitation being implemented by or on behalf of health care providers; thus, solely theoretical papers were excluded; articles addressing aspects of digital facilitation already covered by the included articles identified in stage 1
Article type	Gray literature (ie, literature produced in electronic and print formats outside of commercial publishing), including but not limited to government documents or reports, policy reports, research reports, and working papers	Trial registrations (ie, articles registered on ClinicalTrials.gov or the WHO ICTRP registry)
Article type	Original research	Theoretical and commentary articles and trial registrations (ie, articles registered on ClinicalTrials.gov or the WHO ICTRP registry)

[a]EEA: European Economic Area.

[b]OECD: Organization for Economic Cooperation and Development.

[c]WHO ICTRP: World Health Organization International Clinical Trials Registry Platform.

Data Extraction and Preliminary Analysis

Data from eligible studies were extracted independently by 2 reviewers (EG and SP) using a data-charting form developed for this study. The form was piloted to ensure that data extraction was consistent across reviewers. We extracted data relevant to digital facilitation (digital technology type, facilitation purpose, method, mode of delivery, target population, and setting) and study details (study type, outcomes, size, and setting), aiming to capture health outcomes, staff and patient or caregiver experience, impact on service use (uptake and use of digital services), cost and equity of access to health care services and information, and the nature and extent of other reported outcomes. When considering the outcomes of digital facilitation, we focused on increased uptake and use of web-based services by patients, defining these as indicators of successful facilitation.

Studies were not formally assessed for quality as this was a scoping review, with a great breadth of studies and article types being included. However, reviewers noted the quality of the evidence source, clarity of aims, quality and comprehensiveness of the work, and any conflicts of interest from the authors wherever possible to assist in judging the quality of the overall evidence base for digital facilitation. Given that we did not formally assess the quality of individual studies, we did not report on study quality. Full details of data extraction are available in Multimedia Appendix 2.

Data Analysis, Synthesis, and Typology Development

Preliminary analysis of the data extracted from stage 1 was undertaken to identify gaps in the literature and to inform subsequent stages (Figure 1). Following all extractions, data analysis followed the principles of narrative descriptive synthesis [29]. Key themes were identified and captured during charting, which were then refined and expanded during the preliminary synthesis. The synthesis involved an iterative process of internal study team discussions, analyses, and writing. The typology of digital facilitation approaches was developed through this process of internal team discussion and the synthesis of evidence. Further refinement of themes, initial synthesis, and typology was undertaken through a workshop with study team members, including patient and public involvement representatives.

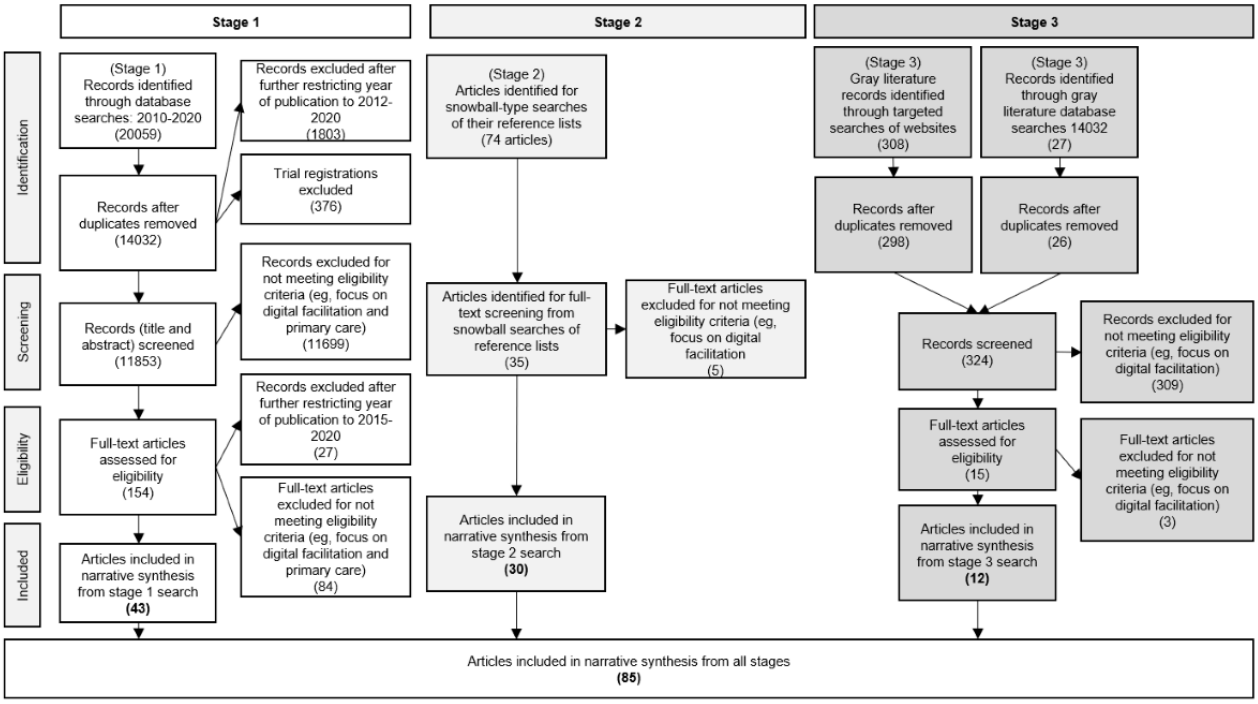
Results

Overview

The PRISMA (Preferred Reporting Item for Systematic Reviews and Meta-Analyses) flow diagram (Figure 2) shows the number of publications retrieved and excluded at each stage. In stage 1, a total of 11,853 records were screened, of which 43 (0.36%) met the inclusion criteria. Later stages identified an additional 42 publications for a total of 85 full-text publications that were included. Multimedia Appendix 3 [6,21,30-105] shows a full list of the included publications and their characteristics.

Publications focused on the United States (30/85, 35%), the United Kingdom (19/85, 22%), other European countries (23/85, 27%), Australia (8/85, 9%), or Canada (1/85, 1%) or adopted an international focus (5/85, 6%). They covered digital facilitation in the primary care sector (48/85, 56%), secondary care sector (5/85, 6%), tertiary care sector (1/85, 1%), or all health care sectors (nonspecific; 31/85, 36%). The publications used various study designs, including quantitative approaches (37/85, 44%; randomized controlled trials [RCTs]: 15/85, 18%; prospective cohort: 4/85, 5%; retrospective cohort design: 1/85, 1%; retrospective observational: 1/85, 1%; longitudinal observational: 1/85, 1%; cross-sectional: 9/85, 11%; pre-post analysis: 1/85, 1%; secondary analysis of data from RCTs: 3/85, 4%), mixed methods approaches (7/85, 8%), qualitative approaches (31/85, 36%), and literature reviews (12/85, 14%). Publications focused on a variety of disease areas, with the most common being diabetes (12/85, 14%) and depression (10/85, 12%).

Figure 2. PRISMA (Preferred Reporting Item For Systematic Review And Meta-Analyses) flow diagram of literature review.



Typology of Digital Facilitation

Overview

A wide variety of digital facilitation efforts were discussed in the literature. In our proposed typology, we categorized them according to whether they were aimed at patients or staff and the purpose of facilitation within them (Table 2). In the following sections, we summarize the descriptive accounts of the different types of digital facilitation and synthesize evidence on whether the approach appeared to be associated with the initial uptake and subsequent use of digital services where available.

Table 2. Typology of digital facilitation approaches.

Typology of digital facilitation	Definition	Examples of facilitation approaches
Digital facilitation aimed at patients		
Promotion	Broad category of digital facilitation that captures ways of raising awareness of and knowledge about digital services, endorsements of specific digital services to patients, and methods of encouraging patients to use them	Recommendation and prescription of digital services and other communication-centered interventions; emails and written reminders; video introductions to digital services
Training and education	Education or training to help patients acquire technical skills to use digital services or to help patients understand what features of a digital service can be most helpful to them	Initial assistance with the use of digital services
Guidance and support	Ongoing help in using digital services provided by clinicians or other primary care staff to patients	Coaching and ongoing guidance from clinicians and other staff
Digital facilitation aimed at primary care staff	Interventions aimed at primary care staff, typically to increase staff's knowledge of digital services so that they can better support patients in their use of the services or to increase their trust in services to increase the likelihood of staff promoting the service to patients	Certified list of apps and websites (to be able to recommend to patients); practice champions (to increase buy-in); training for providers (to generate awareness of web-based services and how to use them)

Digital Facilitation Aimed at Patients

Most (71/85, 84%) articles reported on facilitation efforts aimed directly at patients as distinct from supporting health care staff in helping patients. Facilitation aimed at patients was grouped into three categories: (1) promotion, (2) training, and (3) guidance and support.

Promotion

A lack of knowledge by patients of available web-based services is a significant barrier that primary care staff can help to overcome [30]. The evidence suggests that promotion is a broad category of digital facilitation referring to ways of raising awareness of, as well as knowledge about, digital services; providers endorsing specific digital services to patients; and encouraging patients to use them. Similarly, promotion can take place across a range of media, including on the web, in person

during appointments [31], and in less personalized forms such as placing posters or promotional material in waiting rooms [32].

Examples of web-based promotions include practices featuring links on their websites to promote eConsult (e-consultation and self-help web service [33]), sending reminders or links to web-based services via email or SMS text messages, and using web-based promotional videos. Engaging patients by providing an electronic device such as a tablet for use in the practice waiting room rather than simply relying on verbal recommendations has also been explored in a feasibility study as a way of motivating patients to continue using a web-based self-regulation program once they return home [34]. Verbal recommendation by staff is one of the most widespread, routinely used methods of digital facilitation [35-38].

Training and Education

Training and education may also promote the uptake and use of digital services, both by helping patients acquire technical skills to use web-based services and by helping them understand what features of a web-based service can be most helpful [39]. In the literature we reviewed, training was delivered on the web through videos [40] or offline through in-person support [41,42], presentations, or seminars [43] and was delivered either in a single session [40] or over several sessions [43]. Training and education were commonly combined with helping patients initially sign up for a specific web-based service, such as a patient portal [41,42].

Examples of training included community health care workers conducting home visits to help patients use web-based portals [44]; an in-person tutorial delivered by mental health experts, research assistants, and research nurses within primary care offices [45,106,107]; and a peer support specialist with personal experience in mental illness, substance use, or behavioral concerns to provide technical support to patients using an app within the veteran health system in the United States [46]. There were also fewer resource-intensive facilitation efforts described in the literature, for example, through combinations of written materials, videos, and oral communication about how to use patient portals [41,47].

Training and education delivered through videos were also common [45] and could vary from a short 7-minute video demonstrating how a digital service could be used to a series of 11 videos about how to use patient portals, each with a different theme (eg, patient stories, getting started, signing up and creating a username, accessing different services within the app, and showing patients how to message providers) [106].

Some training and education focused on specific digital services and on providing patients with information on how to use them, whereas other interventions were more general and oriented toward digital literacy and digital health literacy. For example, several qualitative studies recommended that providers suggest computer classes [48], particularly to older adults and patients from minority ethnic groups, to help them use web-based information [49-51]; another study found that digital health literacy courses would be helpful for patients with chronic obstructive pulmonary disease [52].

Guidance and Support

Guidance and support refer to ongoing help provided by clinicians or other primary care staff. It focuses on the technical aspects of using digital services, similar to training, but appears to often focus on interventions that help patients set goals, keep track of progress, and improve adherence and other less technical aspects of digital services. Guidance and support may be particularly important, as a lack of trust and communication has been associated with the discontinued use of web-based platforms [53]. Ongoing guidance and support may be provided through in-person meetings, phone calls, and home visits or in other settings.

Practice champions [108] have been used in primary care to increase the use of web-based services. As experts in a particular web-based service, they provide assistance and ongoing support to patients with the potential to increase both initial uptake and continued use of web-based services thereafter.

Digital Facilitation Aimed at Primary Care Staff

For primary care staff to be able to help patients use web-based services, they must first be aware of what services are available, how they work, why they are useful and trustworthy, and how they can benefit specific patient groups [54]. Health care professionals also need to be clear about their role in terms of endorsing and facilitating web-based services [55]. There is evidence that some GPs are opposed to the use of web-based services by patients [56]. This can include GPs believing that web-based services generate additional workload or preferring to have patients engage directly with the GP [56]. Partly for these reasons, efforts have been made to train primary care staff and increase their knowledge, understanding, and confidence in web-based services.

Some digital facilitation efforts aimed at staff involve interventions to encourage GP practices to adopt more web-based services and actively promote them to their patients. In the United Kingdom, researchers held practice-level discussions with GPs to tackle the strong views held by some GPs against prescribing web-based information, albeit with limited effect [56]. In Spain, an experimental study provided physicians with a list of mobile apps that had been certified by public health authorities and examined the effects of physicians prescribing the apps on patient uptake and use of digital services. As staff buy-in is an important enabler of digital facilitation, having a list of trusted apps can be valuable [57].

Other facilitation efforts focus on training health care practitioners, as studies suggest that staff need training to acquire the necessary technical skills to use web-based services [58] or to effectively reach the target population [59,60]. For example, staff may be trained in communication strategies and relationship building, so that patients or their families are more likely to follow advice to use digital services [59,60]. Such training may be delivered through web-based meetings, face-to-face sessions, and presentations or by sending explanatory videos to the staff [61].

Potential Disadvantages and Risks

Although there is a wealth of literature on the potential harms of digital services, including in terms of health and digital

inequalities, there is less information in the literature included in this review discussing the potential disadvantages of *facilitation efforts* specifically. An example in the included literature was that communication-based facilitation efforts that require high levels of emotional engagement may contribute to distress and fatigue among staff [59]. Another example is that lists of approved apps risk being biased in the considered sample of apps when the onus is on app developers to apply for inclusion in the lists [62]. Some patients have concerns about whether web-based support to encourage continued engagement with digital services would replace valued in-person contact [53]. Email reminders, although sometimes useful, can also cause patients to avoid certain web-based services so as not to receive reminders, although this depends on individual preferences around the frequency of reminders [63]. Finally, the facilitation that provides patients with tablets or computers to use digital services in waiting rooms may compromise patient confidentiality [64].

Evidence also suggests that health care staffs' perceptions of harm from digital services, such as negative impacts on the patient-provider relationships, increased workload, and patients misinterpreting web-based health information, may negatively affect their willingness to recommend digital services to patients [65].

Association of Digital Facilitation With an Increase in Uptake and Use of Digital Services

Overview

The evidence relating to whether different digital facilitation approaches increase the uptake and use of digital services is summarized in Table 3 and is described in the following sections.

Table 3. Evidence on increasing uptake and use of types of digital facilitation approaches.

Typology and digital facilitation effort	Evidence for increasing uptake and use
Promotion	
Recommendation or prescription of digital service to patient	• Staff recommendation or endorsement of a digital service was shown to be one of the most effective ways of increasing patient uptake and use in 2 literature reviews on the topic [21,66]. Qualitative evidence from primary studies also supports staff recommendation or endorsement as an effective way of boosting the use of digital services [35,67,68]. • There is strong evidence from RCTs[a] supporting that prescription and referral pads for digital services are effective in increasing patient uptake [61,69], along with evidence from a review on the topic [70]. • There is some evidence that a list of certified apps and websites (approved by a regulating body) may be effective in enabling providers to prescribe apps and websites to patients [57]. However, when it was implemented by the NHS[b], it had a lack of brand recognition and was ineffective in encouraging the use of high-quality web-based services [71]. • Multiple mixed methods studies suggest that recommendation or endorsement of digital services may be more effective when staff focus on specific aspects of digital service, which will be useful to particular patients, and gradually introduce patients to digital services based on their individual needs at that time [52,65,71-73].
Communication-centered interventions	• Qualitative evidence and evidence from an RCT suggest that recommendation or endorsement of digital services may be more effective when staff are trained in how to best engage patients using specific communication strategies and shared messaging around the service [33,52,53,59,60,64]. For example, these communication strategies can include motivational interviewing and "ICE" formats to address patient ideas, concerns, and expectations. • There is strong evidence from 3 RCTs that interventions that help patients form specific "if-then" plans to use digital services are effective in increasing the continued use of digital services [74].
Email and written reminders	• Mixed methods and qualitative studies have shown that written materials such as brochures, leaflets, and advertisements may be effective in increasing patient use of digital services and are useful in that they require little effort from providers [33,65,75]. • Reminders (eg, SMS text messages and push notifications) have been implemented in some areas [64,76], and feedback from patients and service users suggests that they may be useful in increasing uptake and use [48,53].
Video introductions to digital services	• There is mixed evidence from RCTs on whether video introductions are effective in increasing the uptake of digital services. There is no evidence to support they are effective in increasing the sustained use of digital services [40,47,77,106].
Public information campaigns	• In the United Kingdom, a public information campaign and personalized invitations to invite patients to use an electronic health record system were found to be ineffective in encouraging enrollment [32].
Training	
Initial assistance with use of digital services	• There is mixed evidence from RCTs and quantitative studies on whether initial assistance in registering and logging into digital services is effective in increasing uptake and use [41-43,45]. There is qualitative evidence suggesting patients and providers feel this type of assistance would be useful [78,79]; however, the weight of the evidence suggests that it is likely ineffective and that additional continued support is needed to encourage continued use of digital services. • There is qualitative evidence suggesting that allowing patients to log into and use digital services in primary care practices (eg, in the waiting room on tablets) may be effective in encouraging patients to continue using a service outside of the practice [34,80]. This intervention has been implemented in studies with some success [64].
Technical training support	• There is a body of literature (including strong evidence from a systematic review and an RCT) emphasizing the importance of technical support for using digital services and wider support for digital literacy and digital health literacy in encouraging patient use of digital services, [31,48,51,52,106,109], particularly for older patients, patients from ethnic and racial minority groups, and patients in low-income settings. However, at least one RCT found that simply providing information on using the internet was not effective in increasing the use of digital health services [56].
Guidance and support	

Typology and digital facilitation effort	Evidence for increasing uptake and use
Coaching and ongoing guidance for patients	• There is mixed evidence from RCTs and nonrandomized trials on whether ongoing coaching and support is broadly effective in increasing uptake and sustained use of digital services [44,107,110-112]. The weight of evidence suggests that certain forms of ongoing support are likely effective (see the following sections). • There is strong evidence from 3 RCTs suggesting that ongoing guidance focused on adherence, content of digital services and goal setting are likely more effective than ongoing guidance on only technical aspects of digital services in increasing the use of digital services [74], which is also supported by qualitative evidence [78]. • Qualitative evidence suggests that both face-to-face and telephone support is likely important in encouraging patients to continue to use digital services [53,57,63,109,113].

[a]RCT: randomized controlled trial.

[b]NHS: National Health Service.

Promotion

Recommendation and Prescription of Digital Services and Other Communication-Centered Interventions

Evidence suggests that promotion increases the initial uptake and subsequent use of digital services [66]. A review of promotion methods suggested that endorsement by health care staff is one of the most influential factors affecting patient uptake and use of patient portals [21]. Qualitative findings suggest that recommending digital services to patients is effective in increasing the uptake of those services [35,67,68] where staff focus on specific features of a digital service that will be useful to individual patients [37,52,65], staff are trained in how to best engage patients [60], and staff have a shared understanding of the messaging around digital services [33,64]. Written prescription or referral pads to *prescribe* digital services have also been shown to increase patient uptake of digital services [61,70]. However, a quasi-RCT in the United Kingdom found that providing patients with booklets with general information about using the internet for health purposes was ineffective in increasing their readiness to use electronic health services [56].

Certain communication strategies have been shown to increase the uptake of digital services, such as relationship-building techniques [59]; interviewing and conversational techniques such as motivational interviewing [53]; and discussing patients' ideas, concerns, and expectations to help address patients' misconceptions [52]. Evidence also suggests that gradually introducing patients to digital services, or introducing new features, over the course of several visits rather than all at once can improve the uptake and use of digital services [72,73]. Helping patients form specific plans around the use of digital services was shown to be one of the strongest predictors of adherence in an RCT of an internet-based intervention for depression [74].

Emails and Written Reminders

The written material that health care staff can provide to patients about digital services may be useful in encouraging uptake, incurring minimal time and effort from the staff [65,75]. In a UK study where e-consultation and self-help web services were promoted through posters, leaflets, and advertisements on television screens in waiting rooms and on practice websites, 79% of those who used the web service reported that they discovered the service through these promotion efforts [33]. Reminders for participants can also be helpful in encouraging the uptake and use of digital services [48,53], for example, through SMS text messages sent by receptionists with links to web-based tools [64] or sent to patients at key times, such as when health care staff upload new notes to patient portals, which in one quasi-experimental study resulted in >85% of patients viewing at least one note on the patient portal [76]. However, an RCT examining adherence to an internet-based therapy program for depressive symptoms among high school students found that neither tailored nor standardized emails increased adherence in this group [81].

Training and Education

Initial Assistance With and Education on Use of Digital Services

The evidence is mixed about whether initial assistance with, and education on, the use of digital services increases uptake and continued use. There are contradictory findings on whether initial assistance and education increase the initial uptake or sustained use of digital services.

A quantitative study of the uptake and use of patient portals for patients with chronic kidney disease found that renal clinics that helped patients with initial log-in and registration to the portal had higher levels of portal uptake and use than clinics that did not, with patients 20% more likely to be continued users of the portal after 3 years [42]. A study based on interviews with providers suggests that letting patients use tablet devices or computers in practice waiting rooms may encourage their later use at home [34]. Both health care staff and patients expressed enthusiasm about the potential to access health information [80] and complete digital screening tests [64] on tablets while waiting for appointments.

Although evidence from these studies suggests that the impact of some education and assistance sessions may be long lasting, there is contradicting evidence from other studies indicating that initial training or introductory educational sessions have little impact on use after the initial sign-up [41,43,66]. For example, a study entailed clinical staff providing a 10-minute training session to prospective patient portal users on using and installing a phone app to access the portal, including troubleshooting issues during the training session and providing a pamphlet with further information on the patient portal. It found that although patient interest in the app was high, actual

portal use did not increase after the intervention [41]. Similarly, an RCT regarding the effectiveness of an initial 10-minute standardized personal information session on internet-based depression interventions found that these sessions were ineffective in increasing adherence in an inpatient and outpatient rehabilitation setting for diabetes care [45]. Furthermore, an RCT from the Netherlands showed that initial group education sessions for patients with type 2 diabetes to help them set goals and use web-based platforms were ineffective in increasing the use of the service [43].

Video Introductions to Digital Services

Approximately 5% (4/85) of RCTs evaluated the effect of video introductions on patient uptake and use of digital services, with mixed findings suggesting that video introductions may increase initial uptake but are unlikely to contribute to sustained use. Although 2% (2/85) of studies found that web-based video-based training increased patient *uptake* [47,106], one of the studies found that only 3.5% of patients who were given a video introduction continued to use a portal compared with 1.2% of those who received paper instructions and 0.75% of those who received no intervention, indicating *low sustained use* for all patients [47]. A third RCT found that a 3-minute video was not effective in increasing the uptake or use of a web-based intervention for chronic pain [77]. Another RCT found that a 7-minute video was effective in increasing acceptance of internet-based interventions for depression, although actual use was not measured [40].

Guidance and Support

There is evidence that ongoing support from clinicians and other staff members can increase the use of digital services, although some studies have found these interventions to be ineffective.

Several RCTs have compared the effectiveness of clinicians or other staff in guiding patients in the use of digital services with self-directed services. One of the studies found that patients using web-based therapy for chronic pain who were guided by a psychologist completed more modules than unguided groups and had lower attrition rates [110]. A series of RCTs in Germany comparing different forms of ongoing guidance from clinicians and other staff members assessed how they influenced adherence to digital interventions. The analysis found that both content-focused (personalized written feedback from a psychologist–health coach and reminders to complete modules) and adherence-focused guidance (reminders to complete modules and ability to request feedback from a psychologist–health coach) were equally effective in increasing adherence compared with administrative guidance (technical support in case of computer and internet issues) [82]. However, 2% (2/85) of other RCTs examining the effect of guides on the completion of web-based modules [111] or patient portal use [44] showed either mixed or no evidence for the effectiveness of guides.

Several quantitative studies with nonrandomized control groups also tested the effectiveness of guides in helping patients engage with app content. Some interventions, such as sessions with health coaches [112], hands-on and telephone assistance from nurses, and an intensive course for patients [107], may increase the uptake and use of digital services.

Qualitative evidence also suggests that face-to-face support for patients along with ongoing web support may facilitate the uptake and use of digital services [57,63,109,113]. For example, incorporating digital services into regular care and providing patients with a way of messaging providers for support may encourage sustained engagement [53]. In addition, ongoing training in the use of particular digital services or, more generally, to increase digital literacy skills may encourage uptake and use [48,52,109].

Evidence Relating to Inequality Between Different Population Groups

We found little evidence from studies examining digital facilitation for vulnerable populations, and no studies directly compared different approaches. However, few studies identified strategies that may be effective in increasing the uptake and use of digital services in specific patient populations. For example, a systematic review found that technical training and assistance programs have the best evidence for increasing portal use for vulnerable populations (older adults; racial minorities; and individuals with low socioeconomic status, low health literacy, chronic illness, or disabilities) and that other interventions do not have sufficient evidence [31]. A US study found qualitative evidence that ongoing training, both in the use of a particular service and more generally to increase digital and health literacy skills, can help address the barriers to receiving care faced by African American and Latino patients [50] and patients in low-income areas [51].

There is concern that older people will need extra support to be able to use digital services [66,71]. Some studies showed that older patients were more likely to use digital services after facilitation efforts [106]. Ongoing training and support may also be helpful in encouraging the uptake and use of digital services among older people [48]. Despite concerns about older adult groups being less able or willing to use technology [55,67], evidence suggests that they are nevertheless often willing to use tablets [80], patient portals [21], remote video consultations [73], and health-related apps [62]. Several studies pointed to the importance of ongoing human support [53] and training on both the technical aspects of digital services and general digital literacy skills for older patients [48]. Several studies included subgroup analyses, which revealed that patients with lower health literacy or disabilities were less likely than others to use digital services even after facilitation efforts [43,66,106].

There is some evidence that providers may be more willing and able to engage in digital facilitation efforts with patients who are already confident users of digital services, including the *worried well*, potentially exacerbating inequalities in access to digital health resources [71,83]. A review found that providers are more likely to recommend digital services to patients they perceive as more technologically knowledgeable, and these perceptions may be based on age, socioeconomic status, education level, and ethnic group [36,65].

Factors Affecting Successful Digital Facilitation

Perceptions of Usefulness of the Digital Service

One of the most commonly reported factors influencing the success of digital facilitation efforts in primary (and secondary) care settings is the perception, both from the patients and the health care staff, that digital services will be useful [38,39,55,57,58,72]. Patients are more likely to use services that have been recommended by health care staff if they see the information and functionality as novel [32], if they are able to customize the service to their own needs and preferences [21,63], and if the service is sufficiently specific to fit their needs [84]. Qualitative evidence suggests that the health care staff's likelihood of recommending a digital service to patients may be influenced by the alignment of information within apps and websites with the health information and recommendations that physicians commonly provide to patients [39] and by the availability of evidence that digital services result in patient benefits [30,52].

Time and Capacity in Primary Care

Challenges in terms of staff having sufficient time to implement digital facilitation efforts were commonly identified in the literature [30,34,35,37,46,57,59,64,73,85,113]. The literature also indicated ways of helping to address this issue. Email templates, protocols, and scripts can help staff automate some aspects of patient engagement [70]. Passive facilitation efforts such as posters and brochures can also help mitigate time pressures in primary care [75]. In some studies, it was found to be helpful to have staff other than physicians engage with patients in digital facilitation efforts because of time constraints for physicians [37,64,75,78] or to use the time that patients spend in waiting rooms as an opportunity to facilitate access to web-based services [34,64,80]. Limited time during GP consultations may make it difficult to engage patients in digital facilitation efforts [64]. One of the studies suggested that facilitation efforts may be more feasible during certain kinds of appointments where patients may have less pressing concerns (eg, vaccination-, contraception-, nutritional-, and physical activity–focused appointments) [34].

Buy-in From Health Care Staff

Staff buy-in and motivation were important enablers in many of the studies [42,59,62,86-88], and negative staff attitudes or a lack of motivation toward an intervention were often barriers to facilitation efforts [56,57,60,85]. In several studies, staff buy-in was encouraged through early engagement of staff when developing an intervention, initial education, or training sessions in practices to introduce staff to new web-based services or interventions, ongoing communication with staff, and incorporation of digital services into discussions at staff meetings [30,35,54,60,78]. Ongoing education and training for health care staff on how to use digital services have also been indicated as important in helping them engage in digital facilitation [52,65,70,89].

Reshaping roles in the NHS to incorporate digital services and digital facilitation may also be important in securing staff buy-in [66,73,90]. This not only applies to GPs and nurses but also to wider primary care support teams. Seeing digital facilitation as part of their role rather than something added to their existing job was important in increasing acceptance and buy-in among practice receptionists in a study that required them to send reminders to patients [64].

Patients' and Staff's Trust in, and Knowledge of, Digital Services

Qualitative studies have shown that patients' lack of trust in web-based services can be a barrier to using them [50,80,91], and this is an issue reported by older patients in particular [49]. Fears of loss of confidentiality and security of web-based information may also affect the staff's willingness to recommend digital services to patients [50,65]. Efforts to increase the perceived security of websites were described in the literature, such as the use of third-party seals on patient portal websites [50].

Guidelines and the Role of Regulators

The existence of guidelines that help providers recommend digital services to patients may also be helpful in facilitating efforts [73,87]. Evidence from qualitative studies highlights the importance of simple recruitment criteria, referral guides, and specific triggers that prompt the recommendation of digital services to patients [49,54,60,61,70]. In some cases, mandates for recommending services have also been helpful [89]. In the United Kingdom, it has been suggested that setting targets for GPs to encourage the use of digital services could potentially be effective in increasing patient uptake [66]. Policies that make funding available for training, organizational development, and infrastructure, as well as technology that allows providers to facilitate the use and uptake of digital services, will also be important in increasing use among patients [62,66,73,87,90].

Patients' Trust in Health Care Staff

Trust, perhaps promoted by long-term relationships with health care staff, may be important in patients' use of digital services recommended by those staff [67,114]. Where providers give ongoing support to patients in using a digital service, trusting relationships and a positive, personal tone may boost patients' motivation to participate in digital interventions [53].

Discussion

Typology of Digital Facilitation

We found a rich vein of information about ways in which health care staff in primary care settings can facilitate patients' use of digital services. There is a wide range of approaches to digital facilitation. On the basis of the literature, we developed a novel typology encompassing digital facilitation aimed at patients (promotions, training and education, guidance, and support) and digital facilitation aimed at primary care staff to facilitate patients' use of digital services.

Our review shows diversity in the types of interventions that can be considered under the umbrella term *digital facilitation*. Developing a common framework to define and categorize these advances the evidence base and informs the selection and implementation of different types of facilitation. It also furthers the conceptual understanding of digital facilitation, which is important for informing evaluations of facilitation approaches.

Understanding which approaches work best for whom and in which contexts will be critical for enabling widespread equitable use of digital technologies in primary care, and developing an accurate typology is a necessary first step. We anticipate that future research will seek to refine our proposed typology further. From our review, possible areas for further differentiation may relate to the mode of delivery (eg, in person or on the web) and the degree to which facilitation is passive or active.

Effectiveness of Digital Facilitation Efforts

Our review found evidence that most digital facilitation efforts can support the initial uptake of digital services by patients but that they are unlikely to contribute to sustained use. For example, we found evidence that promotion efforts such as recommendation by practice staff, prescription of digital services, or email and written reminders may increase initial uptake; however, there is little evidence that they lead to sustained use. Similarly, training and education on the use of digital services, such as providing initial assistance with registering for services, also appears to encourage the initial uptake of digital services; however, evidence suggests that these efforts are insufficient to promote long-term use.

Hands-on facilitation approaches, including promotion, guidance, and support by staff, have provided some of the most consistently positive evidence of the usefulness and may be especially important for older adults. This has resource implications as guidance and support take time, and active facilitation takes more time. However, no study has yet examined the cost-effectiveness of digital facilitation. Current mandates in England incentivize the uptake of services and encourage primary care practices to promote web-based services to patients through recommended methods such as posters in physical practices, promotions on practice websites, verbal promotion by practice staff, and promotion via email [3]. However, as identified in our review, promotional approaches may increase initial uptake but seem to not contribute to the sustained use of digital services. It may be that mandates and recommended approaches to digital facilitation need to be revised to recommend approaches such as guidance and coaching, which also incentivize more sustained use. However, without adequate evaluations, including cost-effectiveness studies, it is unclear whether such a mandate is warranted.

Facilitators and Challenges Associated With Digital Facilitation Efforts

Several factors may increase the success of digital facilitation efforts, starting with perceptions among staff and patients that the digital service in question is useful. However, with so many digital services available, it can be challenging for practices to identify appropriate and effective services [115]. Therefore, an important precursor to effective digital facilitation is supporting evaluations of available digital services to help practices and local health authorities understand their impact, affordability, sustainability, and scalability [116]. This would allow practices to prioritize the services likely to be most useful in their local context and has the potential to enhance the trust of health care staff in specific digital services because of their increased knowledge about the services, both of which were found to contribute to the success of facilitation efforts.

Given the current workload pressures on primary care physicians and staff [117-119], actions that reduce the time required to provide facilitation, such as providing guidelines to help practices determine which digital services to prioritize, could increase the success of facilitation efforts. Furthermore, such approaches could also help practices meet the broader aims of digital primary care in the NHS to improve the quality of care to patients and provide efficiency gains to practices [120]. Primary care providers need to feel that the incorporation of patient-facing digital services into their practices is a net gain in terms of workload and efficiency, including any time spent on digital facilitation; otherwise, they may resist efforts to make more services available on the web.

Our review also showed that patients may be more likely to take up and continue to use digital services endorsed or recommended by GPs or other health care staff who they trust. Understanding the value and limitations of these trust relationships is especially important for ensuring the equitable uptake of web-based services. There is substantial evidence that ethnic minority groups in the United Kingdom have lower levels of trust in health care providers and, as a result, face barriers to care [121]. Communication strategies focusing on building trust and positive relationships can help increase the effectiveness of digital facilitation efforts. When positive trust-based relationships exist between GPs and patients, our research suggests that this relationship can be effectively leveraged to help patients access digital services. However, if used in isolation, this approach may leave already disenfranchised groups further excluded from valuable health care resources.

Any implementation of digital facilitation should consider its potential risks and disadvantages, however, this review found little information in the literature specifically discussing the potential disadvantages of facilitation efforts. The available evidence suggests the potential for facilitation efforts to contribute to distress or fatigue among staff. Although this review focused specifically on facilitation efforts, it is reasonable to assume that the harms of digital services being promoted to patients through these facilitation efforts would be important to consider in terms of the risks and disadvantages of facilitation. For example, there are concerns from patients and providers that digital services may replace valued in-person contact or interfere with patient-provider relationships. Holding negative views about web-based services may decrease patients' or providers' willingness to engage in digital facilitation efforts. It would also be important to consider the cost-effectiveness and opportunity costs in terms of primary care staff spending time and resources on digital facilitation rather than other activities.

Agenda for Future Research

Given the push by the NHS for primary care practices to move services to the web [3,5] and the increase in patients' use of web-based services [122], the lack of evidence on how best to facilitate patient access to these services represents a significant gap that should be addressed through future research. A valuable next step would be in-depth qualitative studies that refine our understanding of how digital facilitation occurs in practice,

including identifying where its boundaries lie and how staff and patients engage with facilitation efforts.

Future research should also evaluate the effectiveness and cost-effectiveness of digital facilitation interventions. These studies should focus on outcomes such as the impact on service provision or service use more broadly, as well as on the impacts on patient and staff satisfaction, aspects that were absent in the literature. Future research should also consider the potential unintended impacts of digital facilitation, such as increased inequities in access. Furthermore, limited evidence is available to inform the routine use of digital facilitation in primary care. Consideration of a wider range of outcomes, including patient benefits and costs from service, staff, and patient and caregiver perspectives, will help inform decisions about digital facilitation in primary care practices. Evidence suggests that the context of the patient group, existing relationships, and trust in services can all be important considerations in the effectiveness of facilitation efforts [49,50,53,67,80,91,114]. Understanding this further in the design and evaluation of digital facilitation is important. Our focus on digital facilitation underpinning the organizational aspects of primary care service delivery rather than on exploring the facilitation of digitally delivered therapeutic interventions is a limitation we recognize and which would be a fruitful area for future research.

Only a few studies identified strategies that may be effective in increasing the uptake and use of digital services in specific patient populations. These include technical training and assistance programs for vulnerable populations and providing human support for older adults. Given the existing evidence of inequalities in access to web-based services [15,16] and the role of primary care as the first point of contact for most people [26], this is likely to need careful consideration as NHS England moves forward with its Digital First approach [27]. Evidence on targeting interventions for different groups of patients and for different types of web-based services should be prioritized.

As noted, the COVID-19 pandemic has had an impact on the uptake of digital services [123]; however, it is not clear how digital facilitation efforts have been affected. At the time of our review, no published evidence was available on the impact of COVID-19 on digital facilitation efforts. Timely research is required to more fully understand the pandemic's impact on the provision of digital services in primary care and, crucially, how practices facilitate access, particularly to vulnerable groups and those in most need of support. It is easy to think that digital facilitation may be less important, given that the pandemic has led to a surge in the use of digital services; however, as others argue, ensuring that increases in uptake are sustained will be crucial, and in the context of disrupted and backlogged routine care, digital services are likely to become increasingly important [6].

Strengths and Limitations

As a scoping review, a formal quality assessment of studies was not undertaken, which limited the assessment of the strength of evidence in this review. However, this allowed us to capture the breadth of the literature on digital facilitation in primary care. We were able to describe the breadth of the types of facilitation and provide some assessments of usefulness based on diverse evidence. It is possible that in restricting the selection of publications to 2015 onward, we may have missed earlier publications of relevance; however, from our staged and iterative process in restricting the date, we do not anticipate that this was likely.

Conclusions

The number of drivers to increase the use of digital services in primary care is likely to increase. Digital facilitation is a useful umbrella term that we have introduced into the literature to describe a range of efforts seeking to promote the uptake and use of digital services. Our review found diverse examples of digital facilitation targeting either patients or health care staff. Evidence of its effectiveness was limited, with no evidence of cost-effectiveness. Methods of promotion or initial training appear to be effective in increasing the initial uptake of services but not sustained use without further support. Incentives or requirements for practices to increase the uptake of digital services should also include ongoing use. Despite growing concerns about inequalities in the uptake of digital services, there is limited consideration of this in the literature, either in the design or evaluation of interventions. There is a need to improve both the conceptual understanding and evaluations of digital facilitation. This study offers an initial typology that helps inform both of these key areas of consideration.

Acknowledgments

This study is funded by the National Institute for Health Research (NIHR) [Health Services and Delivery Research Programme 128268]. The views expressed in this publication are those of the author(s) and not necessarily those of the National Institute for Health Research or the Department of Health and Social Care.

EC receives additional support from the National Institute for Health Research Applied Research Collaboration South West Peninsula (NIHR PenARC).

CC has received Honoraria from Bayer and ReCor medical for unrelated projects.

Authors' Contributions

BL contributed to developing the search strategy and typology, evidence synthesis and analysis, drafting the journal article, and reviewing and approving the final article.

SP and EG contributed to developing the search strategy and typology, screening articles, evidence synthesis and analysis, drafting the journal article, and reviewing and approving the final article.
JS led the development of the search strategy and the typology; contributed to evidence synthesis, analysis, and drafting of the journal article; and reviewed and approved the final manuscript.
GA, HA, JC, CC, EC, CM, and EP contributed to developing the search strategy and typology and editing and commenting on the journal article, and reviewed and approved the final article.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Details of literature searches.
[DOCX File , 21 KB - jmir_v24i7e33911_app1.docx]

Multimedia Appendix 2
Data fields in data extraction chart.
[DOCX File , 15 KB - jmir_v24i7e33911_app2.docx]

Multimedia Appendix 3
Full list of publications included scoping review of the literature.
[DOCX File , 143 KB - jmir_v24i7e33911_app3.docx]

References

1. Patient online management information (POMI). NHS England. URL: https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-collections/pomi [accessed 2022-06-03]
2. Mold F, de Lusignan S, Sheikh A, Majeed A, Wyatt JC, Quinn T, et al. Patients' online access to their electronic health records and linked online services: a systematic review in primary care. Br J Gen Pract 2015 Mar;65(632):e141-e151 [FREE Full text] [doi: 10.3399/bjgp15X683941] [Medline: 25733435]
3. 2019/20 General Medical Services (GMS) contract: Guidance and audit requirements for GMS contract. NHS. URL: https://www.england.nhs.uk/publication/2019-20-general-medical-services-gms-contract-guidance-and-audit-requirements-for-gms-contract/ [accessed 2022-06-03]
4. Sheikh A, Anderson M, Albala S, Casadei B, Franklin BD, Richards M, et al. Health information technology and digital innovation for national learning health and care systems. Lancet Digital Health 2021 Jun;3(6):e383-e396. [doi: 10.1016/s2589-7500(21)00005-4]
5. Five year forward view. NHS England. 2014. URL: https://www.england.nhs.uk/five-year-forward-view/ [accessed 2022-06-03]
6. Hutchings R. The Impact of Covid-19 on the Use of Digital Technology in the NHS. London: Nuffield Trust; 2020.
7. Johnsen TM, Norberg BL, Kristiansen E, Zanaboni P, Austad B, Krogh FH, et al. Suitability of video consultations during the COVID-19 pandemic lockdown: cross-sectional survey among Norwegian General Practitioners. J Med Internet Res 2021 Feb 08;23(2):e26433 [FREE Full text] [doi: 10.2196/26433] [Medline: 33465037]
8. Murphy M, Scott LJ, Salisbury C, Turner A, Scott A, Denholm R, et al. Implementation of remote consulting in UK primary care following the COVID-19 pandemic: a mixed-methods longitudinal study. Br J Gen Pract 2021 Jan 17;71(704):e166-e177. [doi: 10.3399/bjgp.2020.0948]
9. Scott A, Bai T, Zhang Y. Association between telehealth use and general practitioner characteristics during COVID-19: findings from a nationally representative survey of Australian doctors. BMJ Open 2021 Mar 24;11(3):e046857 [FREE Full text] [doi: 10.1136/bmjopen-2020-046857] [Medline: 33762248]
10. Solans O, Vidal-Alaball J, Roig Cabo P, Mora N, Coma E, Bonet Simó JM, et al. Characteristics of citizens and their use of teleconsultations in primary care in the Catalan public health system before and during the COVID-19 pandemic: retrospective descriptive cross-sectional study. J Med Internet Res 2021 May 27;23(5):e28629 [FREE Full text] [doi: 10.2196/28629] [Medline: 33970867]
11. Atherton H, Brant H, Ziebland S, Bikker A, Campbell J, Gibson A, et al. Alternatives to the face-to-face consultation in general practice: focused ethnographic case study. Br J Gen Pract 2018 Jan 29;68(669):e293-e300. [doi: 10.3399/bjgp18x694853]
12. Mohammed MA, Montague J, Faisal M, Lamming L. The value of a Patient Access Portal in primary care: a cross-sectional survey of 62,486 registered users in the UK. Univ Access Inf Soc 2019 Oct 22;19(4):855-872. [doi: 10.1007/s10209-019-00693-8]

13. Zanaboni P, Kummervold PE, Sørensen T, Johansen MA. Patient use and experience with online access to electronic health records in Norway: results from an online survey. J Med Internet Res 2020 Feb 07;22(2):e16144 [FREE Full text] [doi: 10.2196/16144] [Medline: 32031538]
14. Marjanovic S, Sim M, Dubow T, Corbett J, Harte E, Parks S, et al. Innovation as a driver of quality and productivity in UK healthcare: creating and connecting receptive places-emerging insights. Rand Health Q 2018 Mar;7(4):1 [FREE Full text] [Medline: 30083413]
15. Atherton H, Ziebland S. What do we need to consider when planning, implementing and researching the use of alternatives to face-to-face consultations in primary healthcare? Digit Health 2016 Nov 21;2:2055207616675559 [FREE Full text] [doi: 10.1177/2055207616675559] [Medline: 29942570]
16. Bryce C, O'Connell MD, Dale J, Underwood M, Atherton H. Online and telephone access to general practice: a cross-sectional patient survey. BJGP Open 2021 Apr 28;5(4):BJGPO.2020.0179. [doi: 10.3399/bjgpo.2020.0179]
17. Facilitating access to online NHS primary care services - current experience and future potential. National Institute for Health Research. 2019. URL: https://fundingawards.nihr.ac.uk/award/NIHR128268 [accessed 2022-06-03]
18. Using online consultations in primary care: implementation toolkit. NHS Digital. URL: https://www.england.nhs.uk/publication/using-online-consultations-in-primary-care-implementation-toolkit/ [accessed 2022-06-03]
19. Huh J, Koola J, Contreras A, Castillo A, Ruiz M, Tedone K, et al. Consumer health informatics adoption among underserved populations: thinking beyond the digital divide. Yearb Med Inform 2018 Aug 29;27(1):146-155 [FREE Full text] [doi: 10.1055/s-0038-1641217] [Medline: 30157518]
20. Donaghy E, Hammersley V, Atherton H, Bikker A, Mcneilly H, Campbell J, et al. Feasibility, acceptability, and content of video consulting in primary care. Br J Gen Pract 2019 Jun 06;69(suppl 1):bjgp19X702941. [doi: 10.3399/bjgp19x702941]
21. 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]
22. 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 Syst 2014 Jan;57:417-427. [doi: 10.1016/j.dss.2012.10.047]
23. Borg K, Boulet M, Smith L, Bragge P. Health Commun 2019 Oct;34(11):1320-1328. [doi: 10.1080/10410236.2018.1485077] [Medline: 29889560]
24. Peters M, Godfrey C, Khalil H, McInerney P, Parker D, Soares C. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015 Sep;13(3):141-146. [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]
25. Leach B, Evangelos G, Parkinson S, Newbould J, Sussex J. Digital facilitation to support patient access to online primary care services. PROSPERO 2020. 2020. URL: https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=189019 [accessed 2022-06-10]
26. What is primary health care? University of Bristol. URL: http://www.bristol.ac.uk/primaryhealthcare/whatisphc.html [accessed 2022-02-08]
27. Digital first primary care. NHS England. URL: https://www.england.nhs.uk/gp/digital-first-primary-care/ [accessed 2022-02-08]
28. Pressures in general practice data analysis. British Medical Association. URL: https://www.bma.org.uk/advice-and-support/nhs-delivery-and-workforce/pressures/pressures-in-general-practice-data-analysis [accessed 2022-06-03]
29. Ryan R. Cochrane Consumers and Communication Review Group: data synthesis and analysis. Cochrane Consumers Commun Rev Group 2013;2013:1-5 [FREE Full text]
30. Nguyen A, Frensham L, Baysari M, Carland J, Day R. Patients' use of mobile health applications: what general practitioners think. Fam Pract 2019 Mar 20;36(2):214-218. [doi: 10.1093/fampra/cmy052] [Medline: 29873708]
31. Grossman L, Masterson Creber RM, Benda N, Wright D, Vawdrey D, Ancker J. 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]
32. Greenhalgh T, Hinder S, Stramer K, Bratan T, Russell J. Adoption, non-adoption, and abandonment of a personal electronic health record: case study of HealthSpace. BMJ 2010 Nov 16;341(nov16 1):c5814 [FREE Full text] [doi: 10.1136/bmj.c5814] [Medline: 21081595]
33. Carter M, Fletcher E, Sansom A, Warren FC, Campbell JL. Feasibility, acceptability and effectiveness of an online alternative to face-to-face consultation in general practice: a mixed-methods study of webGP in six Devon practices. BMJ Open 2018 Feb 15;8(2):e018688 [FREE Full text] [doi: 10.1136/bmjopen-2017-018688] [Medline: 29449293]
34. Plaete J, Crombez G, DeSmet A, Deveugele M, Verloigne M, De Bourdeaudhuij I. What do general practitioners think about an online self-regulation programme for health promotion? Focus group interviews. BMC Fam Pract 2015 Jan 22;16:3 [FREE Full text] [doi: 10.1186/s12875-014-0214-5] [Medline: 25608851]
35. Aarts JW, Faber MJ, Cohlen BJ, Van Oers A, Nelen WL, Kremer JA. Lessons learned from the implementation of an online infertility community into an IVF clinic's daily practice. Hum Fertil (Camb) 2015;18(4):238-247. [doi: 10.3109/14647273.2015.1057901] [Medline: 26167659]
36. Ayre J, Bonner C, Bramwell S, McClelland S, Jayaballa R, Maberly G, et al. Implications for GP endorsement of a diabetes app with patients from culturally diverse backgrounds: a qualitative study. Aust J Prim Health 2020;26(1):52. [doi: 10.1071/py19062]

37. Mares M, Gustafson DH, Glass JE, Quanbeck A, McDowell H, McTavish F, et al. Implementing an mHealth system for substance use disorders in primary care: a mixed methods study of clinicians' initial expectations and first year experiences. BMC Med Inform Decis Mak 2016 Sep 29;16(1):126 [FREE Full text] [doi: 10.1186/s12911-016-0365-5] [Medline: 27687632]
38. Sprenger M, Mettler T, Osma J. Health professionals' perspective on the promotion of e-mental health apps in the context of maternal depression. PLoS One 2017;12(7):e0180867 [FREE Full text] [doi: 10.1371/journal.pone.0180867] [Medline: 28704442]
39. Hjelmager DM, Vinther LD, Poulsen SH, Petersen LS, Jensen MB, Riis A. Requirements for implementing online information material for patients with low back pain in general practice: an interview study. Scand J Prim Health Care 2019 Mar;37(1):60-68 [FREE Full text] [doi: 10.1080/02813432.2019.1569413] [Medline: 30700191]
40. Ebert DD, Berking M, Cuijpers P, Lehr D, Pörtner M, Baumeister H. Increasing the acceptance of internet-based mental health interventions in primary care patients with depressive symptoms. A randomized controlled trial. J Affect Disord 2015 May 01;176:9-17. [doi: 10.1016/j.jad.2015.01.056] [Medline: 25682378]
41. Ramsey A, Lanzo E, Huston-Paterson H, Tomaszewski K, Trent M. Increasing patient portal usage: preliminary outcomes from the MyChart genius project. J Adolesc Health 2018 Jan;62(1):29-35 [FREE Full text] [doi: 10.1016/j.jadohealth.2017.08.029] [Medline: 29169768]
42. Phelps RG, Taylor J, Simpson K, Samuel J, Turner AN. Patients' continuing use of an online health record: a quantitative evaluation of 14,000 patient years of access data. J Med Internet Res 2014 Oct 24;16(10):e241 [FREE Full text] [doi: 10.2196/jmir.3371] [Medline: 25344621]
43. du Pon E, Kleefstra N, Cleveringa F, van Dooren A, Heerdink ER, van Dulmen S. Effects of the proactive interdisciplinary self-management (PRISMA) program on online care platform usage in patients with type 2 diabetes in primary care: a randomized controlled trial. J Diabetes Res 2020;2020:5013142 [FREE Full text] [doi: 10.1155/2020/5013142] [Medline: 32016122]
44. Apter AJ, Localio AR, Morales KH, Han X, Perez L, Mullen AN, et al. Home visits for uncontrolled asthma among low-income adults with patient portal access. J Allergy Clin Immunol 2019 Sep;144(3):846-53.e11 [FREE Full text] [doi: 10.1016/j.jaci.2019.05.030] [Medline: 31181221]
45. Baumeister H, Nowoczin L, Lin J, Seifferth H, Seufert J, Laubner K, et al. Impact of an acceptance facilitating intervention on diabetes patients' acceptance of internet-based interventions for depression: a randomized controlled trial. Diabetes Res Clin Pract 2014 Jul;105(1):30-39. [doi: 10.1016/j.diabres.2014.04.031] [Medline: 24862240]
46. Miller KE, Kuhn E, Yu J, Owen JE, Jaworski BK, Taylor K, et al. Use and perceptions of mobile apps for patients among VA primary care mental and behavioral health providers. Professional Psychol Res Pract 2019 Jun;50(3):204-209. [doi: 10.1037/pro0000229]
47. North F, Hanna BK, Crane SJ, Smith SA, Tulledge-Scheitel SM, Stroebel RJ. Patient portal doldrums: does an exam room promotional video during an office visit increase patient portal registrations and portal use? J Am Med Inform Assoc 2011 Dec;18 Suppl 1:i24-i27 [FREE Full text] [doi: 10.1136/amiajnl-2011-000381] [Medline: 21976028]
48. Nijland N, van Gemert-Pijnen JE, Kelders SM, Brandenburg BJ, Seydel ER. Factors influencing the use of a web-based application for supporting the self-care of patients with type 2 diabetes: a longitudinal study. J Med Internet Res 2011 Sep 30;13(3):e71 [FREE Full text] [doi: 10.2196/jmir.1603] [Medline: 21959968]
49. Moult A, Burroughs H, Kingstone T, Chew-Graham CA. How older adults self-manage distress - does the internet have a role? A qualitative study. BMC Fam Pract 2018 Nov 29;19(1):185 [FREE Full text] [doi: 10.1186/s12875-018-0874-7] [Medline: 30497414]
50. Lyles CR, Allen JY, Poole D, Tieu L, Kanter MH, Garrido T. "I want to keep the personal relationship with my doctor": understanding barriers to portal use among African Americans and Latinos. J Med Internet Res 2016 Oct 03;18(10):e263 [FREE Full text] [doi: 10.2196/jmir.5910] [Medline: 27697748]
51. 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]
52. Slevin P, Kessie T, Cullen J, Butler MW, Donnelly SC, Caulfield B. Exploring the barriers and facilitators for the use of digital health technologies for the management of COPD: a qualitative study of clinician perceptions. QJM 2020 Mar 01;113(3):163-172. [doi: 10.1093/qjmed/hcz241] [Medline: 31545374]
53. van Middelaar T, Beishuizen CR, Guillemont J, Barbera M, Richard E, Moll van Charante EP, HATICE consortium. Engaging older people in an internet platform for cardiovascular risk self-management: a qualitative study among Dutch HATICE participants. BMJ Open 2018 Jan 21;8(1):e019683 [FREE Full text] [doi: 10.1136/bmjopen-2017-019683] [Medline: 29358447]
54. Cowie J, Calveley E, Bowers G, Bowers J. Evaluation of a digital consultation and self-care advice tool in primary care: a multi-methods study. Int J Environ Res Public Health 2018 May 02;15(5):896 [FREE Full text] [doi: 10.3390/ijerph15050896] [Medline: 29724040]
55. Hardiker N, Grant M. Barriers and facilitators that affect public engagement with eHealth services. In: Studies in Health Technology and Informatics. Amsterdam, Netherlands: IOS Press; 2010.

56. Abbott-Garner P, Richardson J, Jones RB. The impact of superfast broadband, tailored booklets for households, and discussions with general practitioners on personal electronic health readiness: cluster factorial quasi-randomized control trial. J Med Internet Res 2019 Mar 11;21(3):e11386 [FREE Full text] [doi: 10.2196/11386] [Medline: 30855234]
57. Lopez Segui F, Pratdepadua Bufill C, Abdon Gimenez N, Martinez Roldan J, Garcia Cuyas F. The prescription of mobile apps by primary care teams: a pilot project in Catalonia. JMIR Mhealth Uhealth 2018 Jun 21;6(6):e10701 [FREE Full text] [doi: 10.2196/10701] [Medline: 29929948]
58. Bernhard G, Mahler C, Seidling HM, Stützle M, Ose D, Baudendistel I, et al. Developing a shared patient-centered, web-based medication platform for type 2 diabetes patients and their health care providers: qualitative study on user requirements. J Med Internet Res 2018 Mar 27;20(3):e105 [FREE Full text] [doi: 10.2196/jmir.8666] [Medline: 29588269]
59. Mahoney N, Gladstone T, DeFrino D, Stinson A, Nidetz J, Canel J, et al. Prevention of adolescent depression in primary care: barriers and relational work solutions. Calif J Health Promot 2017 Aug;15(2):1-12 [FREE Full text] [Medline: 30393470]
60. Molleda L, Bahamon M, St George SM, Perrino T, Estrada Y, Correa Herrera D, et al. Clinic personnel, facilitator, and parent perspectives of eHealth Familias Unidas in primary care. J Pediatr Health Care 2017;31(3):350-361 [FREE Full text] [doi: 10.1016/j.pedhc.2016.11.001] [Medline: 28012799]
61. Byambasuren O, Beller E, Hoffmann T, Glasziou P. mHealth app prescription in Australian general practice: pre-post study. JMIR Mhealth Uhealth 2020 Jun 01;8(6):e16497 [FREE Full text] [doi: 10.2196/16497] [Medline: 32478660]
62. Castle-Clarke S. What Will New Technology Mean for the NHS and Its Patients?: Four Big Technological Trends. London: The Health Foundation, the Institute for Fiscal Studies, The King's Fund and the Nuffield Trust; 2018.
63. van Beurden SB, Simmons SI, Tang JC, Mewse AJ, Abraham C, Greaves CJ. Informing the development of online weight management interventions: a qualitative investigation of primary care patient perceptions. BMC Obes 2018;5:7 [FREE Full text] [doi: 10.1186/s40608-018-0184-6] [Medline: 29449953]
64. 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 24;6(4):e105 [FREE Full text] [doi: 10.2196/mhealth.8778] [Medline: 29691209]
65. Ronda MC, Dijkhorst-Oei L, Vos RC, Rutten GE. Diabetes care providers' opinions and working methods after four years of experience with a diabetes patient web portal; a survey among health care providers in general practices and an outpatient clinic. BMC Fam Pract 2018 Jun 21;19(1):94 [FREE Full text] [doi: 10.1186/s12875-018-0781-y] [Medline: 29929483]
66. Castle-Clarke S, Imison C. The Digital Patient: Transforming Primary Care?. London: Nuffield Trust; 2016.
67. Hoffmann M, Hartmann M, Wensing M, Friederich H, Haun MW. Potential for integrating mental health specialist video consultations in office-based routine primary care: cross-sectional qualitative study among family physicians. J Med Internet Res 2019 Aug 19;21(8):e13382 [FREE Full text] [doi: 10.2196/13382] [Medline: 31429419]
68. Lipschitz J, Miller CJ, Hogan TP, Burdick KE, Lippin-Foster R, Simon SR, et al. Adoption of mobile apps for depression and anxiety: cross-sectional survey study on patient interest and barriers to engagement. JMIR Ment Health 2019 Jan 25;6(1):e11334 [FREE Full text] [doi: 10.2196/11334] [Medline: 30681968]
69. Hassett L, van den Berg M, Lindley RI, Crotty M, McCluskey A, van der Ploeg HP, et al. Digitally enabled aged care and neurological rehabilitation to enhance outcomes with Activity and MObility UsiNg Technology (AMOUNT) in Australia: a randomised controlled trial. PLoS Med 2020 Feb 18;17(2):e1003029 [FREE Full text] [doi: 10.1371/journal.pmed.1003029] [Medline: 32069288]
70. Reynolds J, Griffiths KM, Cunningham JA, Bennett K, Bennett A. Clinical practice models for the use of e-mental health resources in primary health care by health professionals and peer workers: a conceptual framework. JMIR Ment Health 2015;2(1):e6 [FREE Full text] [doi: 10.2196/mental.4200] [Medline: 26543912]
71. Coulter A, Mearns B. Developing care for a changing population: patient engagement and health information technology. Nuffield Trust. 2016. URL: https://www.nuffieldtrust.org.uk/research/developing-care-for-a-changing-population-patient-engagement-and-health-information-technology [accessed 2022-06-04]
72. Longacre M, Grande S, Hager A, Montan M, Bergquist RP, Martensson M, et al. Clinical adoption of mHealth technology to support pediatric cystic fibrosis care in Sweden: qualitative case study. JMIR Pediatr Parent 2018 Dec 10;1(2):e11080 [FREE Full text] [doi: 10.2196/11080] [Medline: 31518297]
73. Rodgers M, Raine G, Thomas S, Harden M, Eastwood A. Informing NHS policy in 'digital-first primary care': a rapid evidence synthesis. Health Serv Deliv Res 2019;7(41). [doi: 10.3310/hsdr07410] [Medline: 31869020]
74. Zarski A, Berking M, Reis D, Lehr D, Buntrock C, Schwarzer R, et al. Turning good intentions into actions by using the health action process approach to predict adherence to internet-based depression prevention: secondary analysis of a randomized controlled trial. J Med Internet Res 2018 Jan 11;20(1):e9 [FREE Full text] [doi: 10.2196/jmir.8814] [Medline: 29326097]
75. Radovic A, Odenthal K, Flores AT, Miller E, Stein BD. Prescribing technology to increase uptake of depression treatment in primary care: a pre-implementation focus group study of SOVA (supporting our valued adolescents). J Clin Psychol Med Settings 2020 Dec;27(4):766-782 [FREE Full text] [doi: 10.1007/s10880-019-09669-5] [Medline: 31630347]

76. Delbanco T, Walker J, Bell SK, Darer JD, Elmore JG, Farag N, et al. Inviting patients to read their doctors' notes: a quasi-experimental study and a look ahead. Ann Intern Med 2012 Oct 02;157(7):461-470 [FREE Full text] [doi: 10.7326/0003-4819-157-7-201210020-00002] [Medline: 23027317]
77. Lin J, Faust B, Ebert DD, Krämer L, Baumeister H. A web-based acceptance-facilitating intervention for identifying patients' acceptance, uptake, and adherence of internet- and mobile-based pain interventions: randomized controlled trial. J Med Internet Res 2018 Aug 21;20(8):e244 [FREE Full text] [doi: 10.2196/jmir.9925]
78. Ford JH, Alagoz E, Dinauer S, Johnson KA, Pe-Romashko K, Gustafson DH. Successful organizational strategies to sustain use of A-CHESS: a mobile intervention for individuals with alcohol use disorders. J Med Internet Res 2015 Aug 18;17(8):e201 [FREE Full text] [doi: 10.2196/jmir.3965] [Medline: 26286257]
79. Sieck CJ, Hefner JL, Schnierle J, Florian H, Agarwal A, Rundell K, et al. The rules of engagement: perspectives on secure messaging from experienced ambulatory patient portal users. JMIR Med Inform 2017 Jul 04;5(3):e13 [FREE Full text] [doi: 10.2196/medinform.7516] [Medline: 28676467]
80. Patel V, Hale TM, Palakodeti S, Kvedar JC, Jethwani K. Prescription tablets in the digital age: a cross-sectional study exploring patient and physician attitudes toward the use of tablets for clinic-based personalized health care information exchange. JMIR Res Protoc 2015 Oct 19;4(4):e116 [FREE Full text] [doi: 10.2196/resprot.3806] [Medline: 26481906]
81. Lillevoll KR, Vangberg HC, Griffiths KM, Waterloo K, Eisemann MR. Uptake and adherence of a self-directed internet-based mental health intervention with tailored e-mail reminders in senior high schools in Norway. BMC Psychiatry 2014 Jan 21;14(1):14 [FREE Full text] [doi: 10.1186/1471-244X-14-14] [Medline: 24443820]
82. Zarski A, Lehr D, Berking M, Riper H, Cuijpers P, Ebert DD. Adherence to internet-based mobile-supported stress management: a pooled analysis of individual participant data from three randomized controlled trials. J Med Internet Res 2016 Jun 29;18(6):e146 [FREE Full text] [doi: 10.2196/jmir.4493] [Medline: 27357528]
83. Cameron G, Alderwick H, Bowers A, Dixon J. Shaping Health Futures: Preparing for Tomorrow's Possibilities Today. London: The Health Foundation; 2019.
84. 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]
85. Careyva B, Shaak K, Mills G, Johnson M, Goodrich S, Stello B, et al. Implementation of technology-based patient engagement strategies within practice-based research networks. J Am Board Fam Med 2016;29(5):581-591 [FREE Full text] [doi: 10.3122/jabfm.2016.05.160044] [Medline: 27613791]
86. O'Cathain A, Drabble SJ, Foster A, Horspool K, Edwards L, Thomas C, et al. Being human: a qualitative interview study exploring why a telehealth intervention for management of chronic conditions had a modest effect. J Med Internet Res 2016 Jun 30;18(6):e163 [FREE Full text] [doi: 10.2196/jmir.5879] [Medline: 27363434]
87. Oliveira HT. Bringing health care to the patient: an overview of the use of telemedicine in OECD countries. OECD. URL: https://www.oecd-ilibrary.org/social-issues-migration-health/bringing-health-care-to-the-patient_8e56ede7-en [accessed 2022-06-04]
88. Weppner WG, Ralston JD, Koepsell TD, Grothaus LC, Reid RJ, Jordan L, et al. Use of a shared medical record with secure messaging by older patients with diabetes. Diabetes Care 2010 Nov;33(11):2314-2319 [FREE Full text] [doi: 10.2337/dc10-1124] [Medline: 20739686]
89. Raza Khan U, Zia TA, Pearce C, Perera K. The MyHealthRecord system impacts on patient workflow in general practices. Stud Health Technol Inform 2019 Aug 08;266:162-167. [doi: 10.3233/SHTI190789] [Medline: 31397318]
90. Rosen R. Delivering General Practice With Too Few GPs. London: Nuffield Trust; 2019.
91. Ricciardi L, Mostashari F, Murphy J, Daniel JG, Siminerio EP. A national action plan to support consumer engagement via e-health. Health Aff (Millwood) 2013 Feb;32(2):376-384. [doi: 10.1377/hlthaff.2012.1216] [Medline: 23381531]
92. Byczkowski TL, Munafo JK, Britto MT. Variation in use of Internet-based patient portals by parents of children with chronic disease. Arch Pediatr Adolesc Med 2011 May;165(5):405-411. [doi: 10.1001/archpediatrics.2011.55] [Medline: 21536954]
93. Mafi JN, Mejilla R, Feldman H, Ngo L, Delbanco T, Darer J, et al. Patients learning to read their doctors' notes: the importance of reminders. J Am Med Inform Assoc 2016 Sep;23(5):951-955 [FREE Full text] [doi: 10.1093/jamia/ocv167] [Medline: 26911830]
94. Masterson Creber RM, Grossman L, Ryan B, Qian M, Polubriaginof F, Restaino S, et al. Engaging hospitalized patients with personalized health information: a randomized trial of an inpatient portal. J Am Med Inform Assoc 2019 Feb 01;26(2):115-123 [FREE Full text] [doi: 10.1093/jamia/ocy146] [Medline: 30534990]
95. Chan B, Lyles C, Kaplan C, Lam R, Karliner L. A comparison of electronic patient-portal use among patients with resident and attending primary care providers. J Gen Intern Med 2018 Dec;33(12):2085-2091 [FREE Full text] [doi: 10.1007/s11606-018-4637-x] [Medline: 30187376]
96. Collins B. Technology and innovation for long-term health conditions. The King's Fund. URL: https://www.kingsfund.org.uk/publications/technology-innovation-long-term-health-conditions [accessed 2022-06-04]
97. Talboom-Kamp EP, Verdijk NA, Kasteleyn MJ, Harmans LM, Talboom IJ, Looijmans-van den Akker I, et al. The effect of integration of self-management web platforms on health status in chronic obstructive pulmonary disease management

in primary care (e-vita study): interrupted time series design. J Med Internet Res 2017 Aug 16;19(8):e291 [FREE Full text] [doi: 10.2196/jmir.8262] [Medline: 28814380]
98. Cafazzo JA, Casselman M, Hamming N, Katzman DK, Palmert MR. Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. J Med Internet Res 2012 May 08;14(3):e70 [FREE Full text] [doi: 10.2196/jmir.2058] [Medline: 22564332]
99. Ray JM, Kemp LL, Hubbard A, Cucciare MA. Developing a peer support protocol for improving veterans' engagement to computer-delivered cognitive behavioural therapy. Behav Cogn Psychother 2017 Mar 22;45(3):253-265. [doi: 10.1017/s1352465816000539]
100. Richardson CR, Buis LR, Janney AW, Goodrich DE, Sen A, Hess ML, et al. An online community improves adherence in an internet-mediated walking program. Part 1: results of a randomized controlled trial. J Med Internet Res 2010 Dec 17;12(4):e71 [FREE Full text] [doi: 10.2196/jmir.1338] [Medline: 21169160]
101. Hale K, Capra S, Bauer J. A framework to assist health professionals in recommending high-quality apps for supporting chronic disease self-management: illustrative assessment of type 2 diabetes apps. JMIR Mhealth Uhealth 2015 Sep 14;3(3):e87 [FREE Full text] [doi: 10.2196/mhealth.4532] [Medline: 26369346]
102. Castle-Clarke S, Kumpunen S, Machaquiero S, Curry N, Imison C. Digital requirements for new primary care models: a briefing for clinicians and managers. Nuffield Trust. 2016. URL: https://www.nuffieldtrust.org.uk/research/digital-requirements-for-new-primary-care-models [accessed 2022-06-04]
103. Honeyman M, Dunn P, McKenna H. A Digital NHS? An Introduction to the Digital Agenda and Plans for Implementation. London: The King's Fund; 2016.
104. Maguire D, Evans H, Honeyman M, Omojomolo D. Digital Change in Health and Social Care. London: The King's Fund; 2018.
105. Yamin CK, Emani S, Williams DH, Lipsitz SR, Karson AS, Wald JS, et al. The digital divide in adoption and use of a personal health record. Arch Intern Med 2011 Mar 28;171(6):568-574. [doi: 10.1001/archinternmed.2011.34] [Medline: 21444847]
106. Lyles CR, Tieu L, Sarkar U, Kiyoi S, Sadasivaiah S, Hoskote M, et al. A randomized trial to train vulnerable primary care patients to use a patient portal. J Am Board Fam Med 2019 Mar 08;32(2):248-258 [FREE Full text] [doi: 10.3122/jabfm.2019.02.180263] [Medline: 30850461]
107. Talboom-Kamp EP, Verdijk NA, Kasteleyn MJ, Harmans LM, Talboom IJ, Numans ME, et al. High level of integration in integrated disease management leads to higher usage in the e-vita study: self-management of chronic obstructive pulmonary disease with web-based platforms in a parallel cohort design. J Med Internet Res 2017 May 31;19(5):e185 [FREE Full text] [doi: 10.2196/jmir.7037] [Medline: 28566268]
108. Shaw EK, Howard J, West DR, Crabtree BF, Nease DE, Tutt B, et al. The role of the champion in primary care change efforts: from the State Networks of Colorado Ambulatory Practices and Partners (SNOCAP). J Am Board Fam Med 2012 Sep 05;25(5):676-685 [FREE Full text] [doi: 10.3122/jabfm.2012.05.110281] [Medline: 22956703]
109. Abidi S, Vallis M, Piccinini-Vallis H, Imran SA, Abidi SS. Diabetes-related behavior change knowledge transfer to primary care practitioners and patients: implementation and evaluation of a digital health platform. JMIR Med Inform 2018 Apr 18;6(2):e25 [FREE Full text] [doi: 10.2196/medinform.9629] [Medline: 29669705]
110. Lin J, Paganini S, Sander L, Lüking M, Ebert DD, Buhrman M, et al. An internet-based intervention for chronic pain. Dtsch Arztebl Int 2017 Oct 13;114(41):681-688 [FREE Full text] [doi: 10.3238/arztebl.2017.0681] [Medline: 29082858]
111. 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]
112. Possemato K, Kuhn E, Johnson E, Hoffman JE, Owen JE, Kanuri N, et al. Using PTSD Coach in primary care with and without clinician support: a pilot randomized controlled trial. Gen Hosp Psychiatry 2016;38:94-98. [doi: 10.1016/j.genhosppsych.2015.09.005] [Medline: 26589765]
113. Littlewood E, Duarte A, Hewitt C, Knowles S, Palmer S, Walker S, REEACT Team. A randomised controlled trial of computerised cognitive behaviour therapy for the treatment of depression in primary care: the randomised evaluation of the effectiveness and acceptability of computerised therapy (REEACT) trial. Health Technol Assess 2015 Dec;19(101):viii, xxi-viii, 171 [FREE Full text] [doi: 10.3310/hta191010] [Medline: 26685904]
114. Lyles CR, Sarkar U, Ralston JD, Adler N, Schillinger D, Moffet HH, et al. Patient-provider communication and trust in relation to use of an online patient portal among diabetes patients: the diabetes and aging study. J Am Med Inform Assoc 2013;20(6):1128-1131 [FREE Full text] [doi: 10.1136/amiajnl-2012-001567] [Medline: 23676243]
115. Shemtob L, Littlewood R. Challenges for digital services in the NHS: drowning in a sea of innovation. Br J Gen Pract 2019 Jun 27;69(684):326-327. [doi: 10.3399/bjgp19x704177]
116. Oliver D. David Oliver: lessons from the Babylon Health saga. BMJ 2019 Jun 05;365:l2387. [doi: 10.1136/bmj.l2387] [Medline: 31167923]
117. Croxson CH, Ashdown HF, Hobbs FR. GPs' perceptions of workload in England: a qualitative interview study. Br J Gen Pract 2017 Jan 16;67(655):e138-e147. [doi: 10.3399/bjgp17x688849]

118. Hobbs FD, Bankhead C, Mukhtar T, Stevens S, Perera-Salazar R, Holt T, et al. Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007–14. Lancet 2016 Jun;387(10035):2323-2330. [doi: 10.1016/s0140-6736(16)00620-6]
119. Roland M, Everington S. Tackling the crisis in general practice. BMJ 2016 Feb 17;352:i942. [doi: 10.1136/bmj.i942] [Medline: 26887896]
120. Digital primary care. NHS England. URL: https://www.england.nhs.uk/digitaltechnology/digital-primary-care/ [accessed 2021-08-20]
121. Kapadia D, Zhang J, Salway S, Nazroo J, Booth A, Villarroel-Williams N. Ethnic inequalities in healthcare: a rapid evidence review. NHS Race & Health Observatory. URL: https://www.nhsrho.org/publications/ethnic-inequalities-in-healthcare-a-rapid-evidence-review/ [accessed 2022-06-04]
122. Dramatic annual surge in online GP services as patients sign up for convenience. NHS England. URL: https://www.england.nhs.uk/2018/03/dramatic-annual-surge-in-online-gp-services-as-patients-sign-up-for-convenience/ [accessed 2022-04-05]
123. Locked out: digitally excluded people's experiences of remote GP appointments. Healthwatch. URL: https://www.healthwatch.co.uk/report/2021-06-16/locked-out-digitally-excluded-peoples-experiences-remote-gp-appointments [accessed 2022-06-04]

Abbreviations

GP: general practitioner
NHS: National Health Service
PRISMA: Preferred Reporting Item For Systematic Reviews And Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized controlled trial

Edited by A Mavragani; submitted 29.09.21; peer-reviewed by L Shang, M Carter; comments to author 21.12.21; revised version received 15.02.22; accepted 10.04.22; published 14.07.22.

Please cite as:
Leach B, Parkinson S, Gkousis E, Abel G, Atherton H, Campbell J, Clark C, Cockcroft E, Marriott C, Pitchforth E, Sussex J
Digital Facilitation to Support Patient Access to Web-Based Primary Care Services: Scoping Literature Review
J Med Internet Res 2022;24(7):e33911
URL: https://www.jmir.org/2022/7/e33911
doi:10.2196/33911
PMID:35834301

Review

Using an Integrated Framework to Investigate the Facilitators and Barriers of Health Information Technology Implementation in Noncommunicable Disease Management: Systematic Review

Meekang Sung[1]; Jinyu He[2], BA; Qi Zhou[3], MMed; Yaolong Chen[3], MD; John S Ji[2], ScD; Haotian Chen[2], BA; Zhihui Li[2,4], ScD

[1]College of Pharmacy, Seoul National University, Seoul, Republic of Korea

[2]Vanke School of Public Health, Tsinghua University, Beijing, China

[3]Evidence-Based Medicine Center, School of Basic Medical Sciences, Lanzhou University, Lanzhou, China

[4]Institute for Healthy China, Tsinghua Universtiy, Beijing, China

Corresponding Author:
Zhihui Li, ScD
Vanke School of Public Health
Tsinghua University
Haidian District
Beijing, 10084
China
Phone: 86 13718264652
Email: zhihuili@mail.tsinghua.edu.cn

Abstract

Background: Noncommunicable disease (NCD) management is critical for reducing attributable health burdens. Although health information technology (HIT) is a crucial strategy to improve chronic disease management, many health care systems have failed in implementing HIT. There has been a lack of research on the implementation process of HIT for chronic disease management.

Objective: We aimed to identify the barriers and facilitators of HIT implementation, analyze how these factors influence the implementation process, and identify key areas for future action. We will develop a framework for understanding implementation determinants to synthesize available evidence.

Methods: We conducted a systematic review to understand the barriers and facilitators of the implementation process. We searched MEDLINE, Cochrane, Embase, Scopus, and CINAHL for studies published between database inception and May 5, 2022. Original studies involving HIT-related interventions for NCD management published in peer-reviewed journals were included. Studies that did not discuss relevant outcome measures or did not have direct contact with or observation of stakeholders were excluded. The analysis was conducted in 2 parts. In part 1, we analyzed how the intrinsic attributes of HIT interventions affect the successfulness of implementation by using the intervention domain of the Consolidated Framework for Implementation Research (CFIR). In part 2, we focused on the extrinsic factors of HIT using an integrated framework, which was developed based on the CFIR and the levels of change framework by Ferlie and Shortell.

Results: We identified 51 papers with qualitative, mixed-method, and cross-sectional methodologies. Included studies were heterogeneous regarding disease populations and HIT interventions. In part 1, having a relative advantage over existing health care systems was the most prominent intrinsic facilitator (eg, convenience, improvement in quality of care, and increase in access). Poor usability was the most noted intrinsic barrier of HIT. In part 2, we mapped the various factors of implementation to the integrated framework (the coordinates are shown as *level of change-CFIR*). The key barriers to the extrinsic factors of HIT included health literacy and lack of digital skills (*individual-characteristics of individuals*). The key facilitators included physicians' suggestions, cooperation (*interpersonal-process*), integration into a workflow, and adequate management of data (*organizational-inner setting*). The importance of health data security was identified. Self-efficacy issues of patients and organizational readiness for implementation were highlighted.

Conclusions: Internal factors of HIT and external human factors of implementation interplay in HIT implementation for chronic disease management. Strategies for improvement include ensuring HIT has a relative advantage over existing health care; tackling usability issues; and addressing underlying socioeconomic, interpersonal, and organizational conditions. Further research should

focus on studying various stakeholders, such as service providers and administrative workforces; various disease populations, such as those with obesity and mental diseases; and various countries, including low- and middle-income countries.

(J Med Internet Res 2022;24(7):e37338) doi:10.2196/37338

KEYWORDS

health information technology; noncommunicable disease management; chronic disease management; systematic review; implementation science

Introduction

Background

Noncommunicable diseases (NCDs) are the number one cause of death and disability in the world [1]. According to World Health Organization (WHO) estimates, NCDs caused around 1.6 million disease-adjusted life years worldwide in 2019, accounting for 62% of the total disease-adjusted life years [2]. To lessen the impact of NCDs on individuals and the society, investing in better management is critical [3]. However, effective management of NCDs has many challenges, including fragmented health systems, difficulties in information exchange, and a lack of interoperable clinical information systems [4].

Health information technology (HIT) has been highlighted to overcome these barriers. HIT refers to the electronic system used to store, share, and analyze health information. This includes, but is not limited to, electronic health records (EHRs), personal health records, and electronic prescribing [5]. HIT could improve the quality of care by reducing paperwork, reducing medical errors, minimizing repetitive medical tests, enabling the collaboration of medical professionals over long distances, and reducing the cost of treatment of chronically ill patients [6]. In addition, HIT can increase patients' empowerment by helping them develop self-awareness of NCDs [7,8].

Various health care systems have implemented HIT. In 2017, 94% of hospitals in the United States were using EHR systems for managing clinical data [9]. However, many low- and middle-income countries (LMICs) are not quite finished with adapting HIT [10]. For example, EHR systems are not properly used in more than 50% of developing countries [11,12]. This failure is due to resistance and opposition to changing to electronic systems [13], lack of organizational readiness [14], or lack of funding and lack of technical and computer skills of personnel [15]. Developed countries are also heading toward the adaptation of next-generation HITs [16], such as personal health records, patient-centered care, multi-disciplinary care, health information exchange, and integration of artificial intelligence into the health care system. In any case, implementing HIT is challenging, and thus, it is critical to analyze the barriers and facilitators of HIT implementation.

Prior Work

Implementation of HIT is affected by both the inherent characteristics of HIT (eg, the novelty of the technology and advantages HIT gives to users) and the external factors of HIT (eg, perceptions and behaviors related stakeholders have about implementing new technology). Some studies explored the challenges in a general context, where design and usability issues were mentioned [17-19]. These studies have limitations in understanding the perspectives of various stakeholders. Other previous research concentrated on a specific topic, such as diabetes management [20-25] or one type of HIT (eg, patient web portal) [22], which is insufficient for understanding HIT implementation in a more general setting. Frameworks have helped understand the implementation processes of various topics. For example, Webb et al [8] integrated the level theory by Ferlie and Shortell to understand perinatal mental health care, and Esponda et al used the Consolidated Framework for Implementation Research (CFIR) [26] to analyze mental health implementation [27]. However, determinant frameworks have been used scarcely in understanding HIT implementation. The existing frameworks also have limitations in differentiating between whether a factor is an intrinsic characteristic of HIT or a human factor related to the stakeholders.

The Goal of This Study

Therefore, our objective was to tackle the research gap regarding the implementation of HIT for chronic disease management. We specifically aimed to identify the barriers and facilitators, analyze how these factors influence the HIT implementation process, and identify key areas for future action. We will develop a framework for understanding implementation determinants to synthesize available evidence.

Methods

Search Strategy and Selection Criteria

In this systematic review, literature searches and study selection followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines [28] (Multimedia Appendix 1). As the review did not evaluate a direct health-related outcome, it did not meet the criteria for registration of the protocol with PROSPERO. The author MS searched the MEDLINE, Cochrane, Embase, CINAHL, and Scopus databases for research articles published between database inception and May 5, 2022.

Boolean operators were used to combine relevant search terms related to NCDs (eg, "noncommunicable diseases," "chronic diseases," "diabetes," and "hypertension"), HIT (eg, "health information technology," "electronic health records," "personal health records," and "electronic prescribing"), and implementation outcomes (eg, "barrier" and "facilitator"). Based on the definition of HIT [5], search phrases for HIT also included a wide range of HIT-related literature.

The search syntax was devised and written by MS and reviewed by ZL. The full search syntax can be found in Multimedia Appendix 2. The initial search was completed on August 11, 2021. Forward and backward searches of included studies were

completed by October 31, 2021. The supplementary search was completed by May 5, 2022.

Studies were eligible if they involved HIT-related interventions (eg, EHRs, personal health records, and electronic prescribing), involved interventions that were used for NCD management, and examined implementation outcomes (ie, barriers or facilitators). Studies were included if they were published in peer-reviewed academic journals and had direct contact with or direct observation of different stakeholders, such as patients, the public (consumers), companies, and health professionals. The articles included were required to have full text available and be written in English.

Studies were excluded if they were not related to chronic disease management, did not implement HIT-related interventions (eg, studies that concentrated on digital health interventions that were not related to HIT), had an outcome that was not focused on implementation, or did not discuss facilitators and barriers (eg, studies that reviewed the effectiveness of HIT).

Study Selection

Search results were imported into EndNote 20 (Clarivate). After removing duplicates, MS and JH independently double-screened all titles and abstracts. The interrater reliability between the first and second screeners was 58% in the first screening. Both authors discussed all disagreements and were able to agree on all selections of papers (κ=100%). The full texts of the included papers were then assessed for eligibility by MS and JH. The interrater reliability (κ) was 71% in the initial selection of full-text papers. Both authors discussed all disagreements and came to an agreement on all included studies. If necessary, a third author (ZL) mediated agreement.

Data Collection and Data Items

Extraction of data on author, year, country, study design, data collection methods, participants, intervention stage, target population, HIT program/intervention, and addressed stakeholders was performed by MS and JH into an Excel spreadsheet (Microsoft Corp). The full texts of the studies were also extracted to NVivo (Release 1.5) software (QSR International), which allows for line-by-line coding. Each paper was read in full, and relevant parts of the text were applied to the relevant code. Data extraction followed the data extraction form (Multimedia Appendix 3), which was guided by the Cochrane Systematic Review for Intervention Data Collection form [29].

Critical Appraisal of Studies

MS and JH independently conducted quality assessments of the included studies using several appraisal tools based on the type of research. Joanna Briggs Critical Appraisal Tools were used for qualitative research [30], the Mixed Methods Appraisal Tool [31] was used for mixed methods studies, and the Center for Evidence-Based Management Critical Appraisal Checklist was used for cross-sectional studies [32]. Multimedia Appendix 4 explains each quality appraisal method in detail. Each point of the Joanna Briggs Critical Appraisal Tools can be coded into either yes, no, unclear, or not applicable. Each point of the Mixed Methods Appraisal Tool and the Center for Evidence-Based Management Critical Appraisal Checklist can be coded into yes, no, or cannot tell. Where most questions within a domain or a paper were answered with yes, it was rated as having high quality, and where the majority were answered with no, it was rated as having low quality. Medium quality was when there was a mixture of yes and no answers. The note in Multimedia Appendix 5 explains the detailed criteria for high, medium, and low quality for each type of research. Studies were not excluded based on quality to capture as much literature as possible, but low-quality studies were not used to draw conclusions.

Synthesis of Results

Enhancing Transparency in Reporting the Synthesis of Qualitative Research guidelines were followed (Multimedia Appendix 6) [33]. We used the best-fit framework synthesis approach [34]. First, statements referring to facilitators or barriers of the implementation of HIT-related interventions were extracted line by line. Second, full texts of studies were exported to NVivo for analysis. Statements referring to facilitators or barriers of the implementation of HIT-related interventions were extracted line by line and coded. Third, codes were reread and assigned a descriptive theme based on their content. Once all codes were assigned, various implementation frameworks were assessed for their fit with the existing frameworks (eg, CFIR [26], Reach Effectiveness Adoption Implementation Maintenance [35], socioecological model [36], and levels of change framework by Ferlie and Shortell [37]) to structure themes. The CFIR and the levels of change framework were selected since they best matched the codes and descriptive themes that were derived in this review.

Our analysis was conducted in two parts. Figure 1 illustrates the study design. In part 1, we aimed to understand the inherent characteristics of HIT that affect implementation. The intervention domain ("characteristics of the intervention implemented") of the CFIR was found to fit best and was therefore used. The CFIR, which has been extensively used in research, has a comprehensive categorization of implementation determinants informed by both empirical findings and theory. It is composed of the following 5 domains: (1) intervention, (2) outer setting, (3) inner setting, (4) individuals, and (5) process. The intervention domain is constructed of 8 subconstructs, which help analyze the complex and multi-faceted characteristics of HIT. Among the 8 subconstructs (innovation source, evidence strength and quality, relative advantage, adaptability, trialability, complexity, design quality and usability, cost), "innovation source" and "trialability" did not have matching concepts in our findings and were therefore excluded.

In part 2, we conducted a stakeholder analysis with the integrated framework. The integrated framework was developed based on the CFIR and the levels of change framework by Ferlie and Shortell, as shown in Figure 2. The latter 4 domains of the CFIR involved various stakeholders and their relations. However, the CFIR is limited in identifying which specific stakeholders are involved with a factor.

Figure 1. Study design. CFIR: Consolidated Framework for Implementation Research; HCP: health care provider.

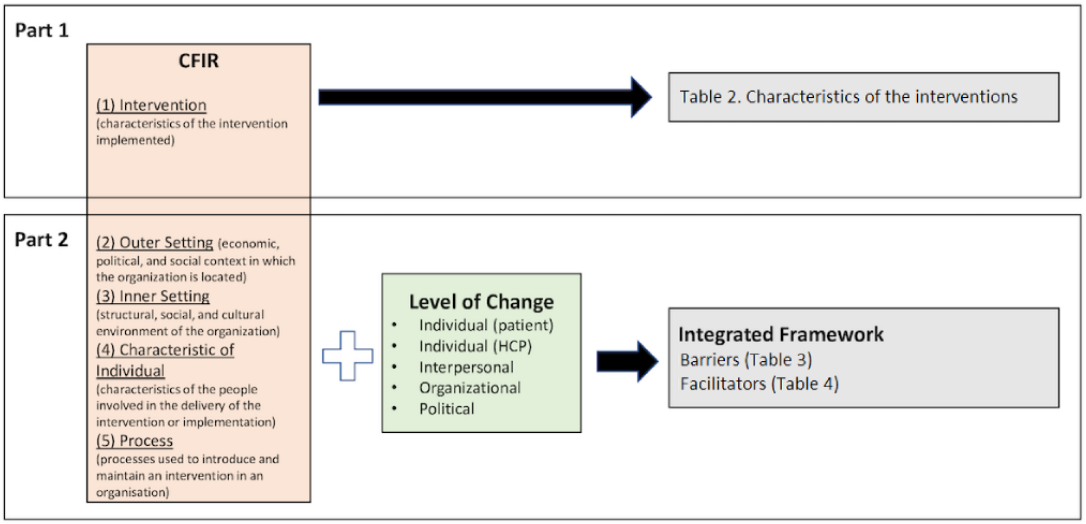
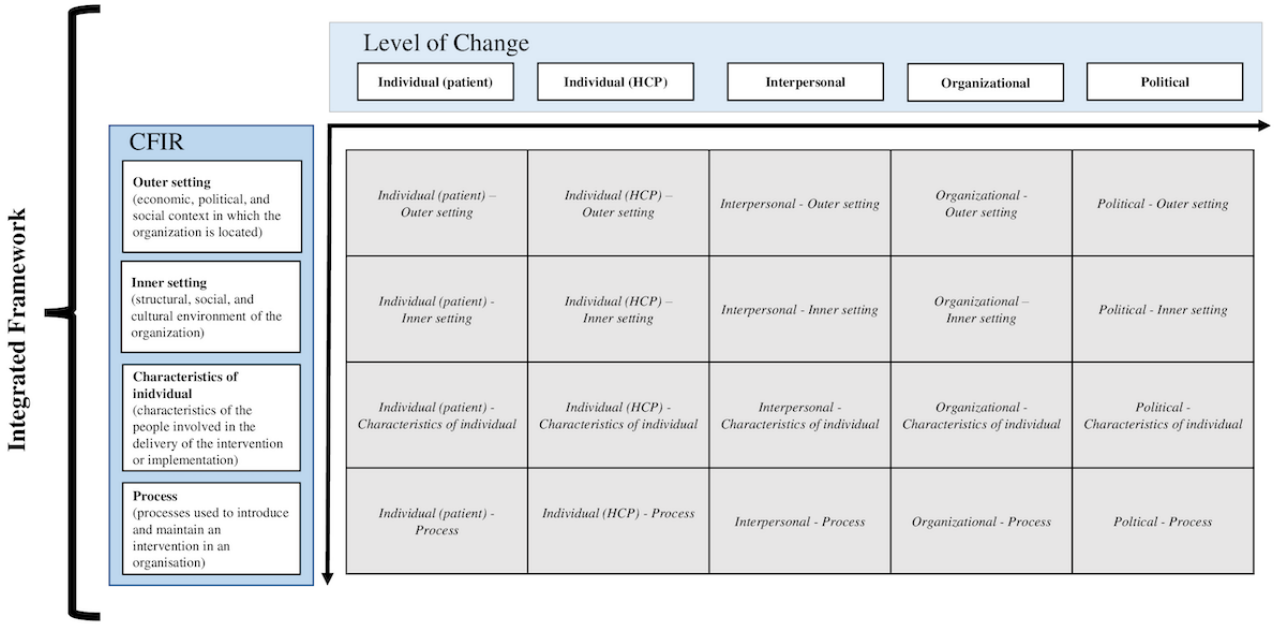


Figure 2. Diagram of the integrated framework. CFIR: Consolidated Framework for Implementation Research; HCP: health care provider.



The levels of change framework, which is also frequently used in the literature, categorizes factors on the following 4 levels: (1) individual, (2) care team, (3) organizational structure, and (4) the wider environment [37]. This framework compensates for the CFIR because it can identify which stakeholders are involved in a factor. Moreover, it can explain at which level the factors are being affected. However, since it is only constructed of 4 levels, it fails to deliver a specific view and separately categorize disparate factors.

By combining the CFIR and the levels of change framework, we could complement each framework's weaknesses. We first modified the categories of the levels of change framework as individual factors (patients and health care providers [HCPs]), interpersonal factors, organizational factors, and political factors. Then, we combined the 2 frameworks to develop a novel integrated framework. Themes that could not be explained by the original frameworks were identified and synthesized into the integrated framework. After developing the integrated framework, codes were reread and assigned deductively. Data coding was undertaken with NVivo (Release 1.5) software.

We placed the CFIR constructs on the vertical axis and the level of change categories on the horizontal axis and mapped relevant factors of implementation in matching coordinates (Figure 2). A factor showing "individual" on the horizontal axis and "outer setting" on the vertical axis, for example, acts at the individual level and is related to the outer setting of implementation. The most mentioned *level of change-CFIR* sections are explained in detail in the Results.

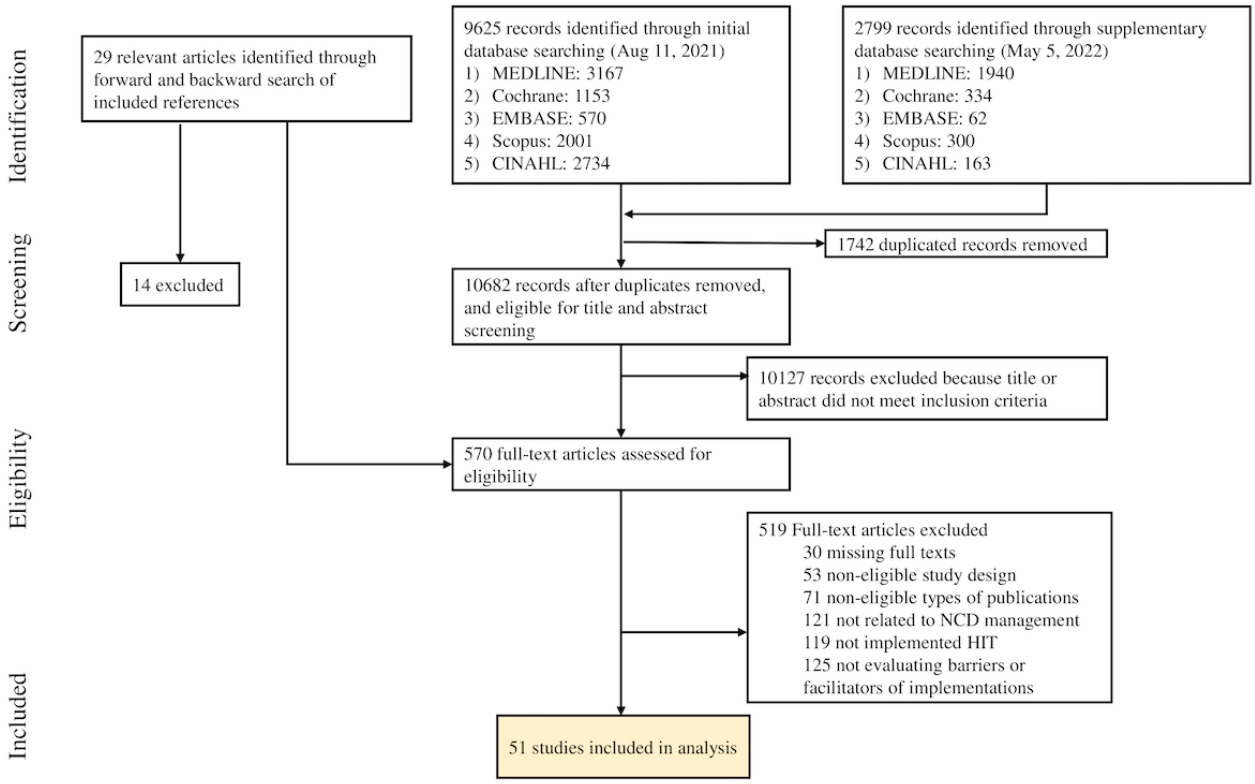
This method helps to understand the overall picture because it provides the location (horizontal and vertical) of factors, and the categories are more specified than either the CFIR or the levels of change framework.

Results

Study Selection

We identified 12,424 records through database searches (Figure 3). A total of 9625 articles were from the initial search, and additional 2799 articles were added from the supplementary search. After removing duplicates, 10,682 citations were left. During the full-text screening of 555 articles, 29 articles identified by the forward and backward searches of the included references were further screened for eligibility, of which 15 articles were finally added. Of a total of 10,697 articles, 570 were identified as potentially relevant records after screening the titles and abstracts. After full-text screening, 51 studies were included for analysis. Figure 3 describes the number of papers excluded for each exclusion criteria.

Figure 3. Study selection. HIT: health information technology; NCD: noncommunicable disease.



Included studies were heterogeneous, with different sample sizes, interventions being implemented, countries of origin, and methodologies. Programs used qualitative, mixed-method, or cross-sectional study designs. A total of 34 studies were qualitative [38-71]. Common qualitative methods for data collection included in-depth interviews and focus groups. The sample sizes of qualitative studies ranged from 18 to 110. Twelve studies used a mixed methods design [72-84]. Common methods for data collection were surveys, questionnaires, or descriptive statistics mixed with qualitative studies. Four studies used quantitative methodology, 3 used cross-sectional survey methodology [85-87], and 1 extracted data from an electronic medical record system [88].

MS and JH independently completed the assessment for the included papers. The appraisal of quality was the same for 37 (73%) of the 51 papers. All disagreements were discussed by SM and HJ, and if necessary, a third author (ZL) mediated agreement. The final appraisal was based on agreed answers. Of the 51 papers, 31 were determined as high quality, 18 as medium quality, and 2 as low quality. The detailed quality evaluation by quality appraisal domains is shown in Multimedia Appendix 5.

Detailed characteristics of the included 51 studies can be found in Multimedia Appendix 7. Most (30/51) of the included studies addressed diabetes [41,42,45,47,49,51-54,56,57,59,60,64,65,72-77,79-82,85,86,88]. Other target populations addressed were as follows: cancer [40,44,66,68,71,84,87], general primary care [46,50,58,78,83], multiple chronic conditions [38,39,63,70], hypertension [57,73,79], mental health [54,55,60], general health care [61], cardiovascular diseases [69], heart disease [60], hyperlipidemia [79], elderly and disabled [43], and chronic kidney disease [62].

Table 1 presents the characteristics of the included studies by type of HIT intervention, target population, country, and stakeholder. The most reported types of HIT interventions were patient portals [46,49,61,63,74,75,81,83,85,86], electronic health registries [54,57,59,62,66,70,71,84,87,88], clinical decision support systems [50,51,55,64,65,69,73,76,78], personal health records [38,39,42-44,56,58,68,77], integrative care modules [45,60,77,79,82], patient decision aids [47,48], digital education programs [41], self-management programs [80], shared decision-making [53], tailored messages [72], general HIT [67], and other programs [40,52]. Most studies primarily focused on the factors that affect patients or HCPs. Some literature reported other stakeholders, such as information technology employees [61], family [44], caregivers [46,83], vendors [59], care

managers [48,61], educators [52], and staff (ie, nurse practitioners and physician assistants) [76].

Table 1. Characteristics of the included studies.

Characteristic	Value (N=51), n[a]
Type of HIT[b] intervention	
Patient portals	10
Electronic health registries	10
Personal health records	9
Clinical decision support systems	9
Integrative care modules	4
Patient decision aids	2
Other HIT-based management	2
Digital education programs	1
Self-management programs	1
Shared decision-making	1
Tailored messages	1
General HIT	1
Target population	
Diabetes	30
Cancer	7
General primary care	5
Multiple chronic conditions	4
Hypertension	3
Mental health	3
Heart disease	1
Hyperlipidemia	1
Elderly and disabled	1
Chronic kidney disease	1
Country	
United States	30
The Netherlands	4
Canada	4
Australia	2
Malaysia	2
Malawi	2
United Kingdom	1
Scotland	1
Brazil	1
Finland	1
Germany	1
Iran	1
Uganda	1
Stakeholder	
Patients	37
Health care providers	27

Characteristic	Value (N=51), n[a]
Vendors	8
Staff/clinic manager	5
Caregivers	2
Information technology employee	1
Researcher	1

[a]Number of included studies.

[b]HIT: health information technology.

Part 1: Inherent Characteristics of HIT Interventions

We coded the inherent characteristics of HIT implementation into barriers and facilitators (Table 2). Detailed definitions and reflective quotes can be found in Multimedia Appendix 8.

Evidence strength and quality was both a facilitator and barrier. A trustworthy knowledge base, such as reliable data sets and recommendations from trusted peers, facilitated HIT use [50]. However, stakeholders would be reluctant in adapting HIT if they did not trust the technology [39,57,75]. For instance, some providers perceived patient-recorded data as unreliable and therefore had a lack of desire to use patient portals [39].

Table 2. Inherent characteristics of health information technology interventions as barriers and facilitators.

Characteristic	Barriers	Facilitators
Evidence strength and quality	Unreliability of data [39,57,75] (3 mentions)	Ensuring reliability [25,50,57] (3 mentions)
Relative advantage	Threaten the HCP[a]-patient relationship [49,50], reduce the quality of care [49], unhelpful [49,51,72,76], and provoke negative emotions [38,39] (9 mentions)	Convenience [42,46,49,52,75,81], help HCP-patient communication [46,63-65,70,76,87,89], help monitoring [52], engagement [42,76,82], improve disease management [46,49,56], improve data quality [71,87], improve quality of care [45,46,49,75,80,90], improve awareness [40,49,56,63,76,77], efficiency [63,71,87], increase access [42,45,49,75,77,85,90], perceived usefulness [44,49,51,61,67,69,75,85], and reduce risk of error [87] (54 mentions)
Adaptability	Inapplicability [50,58,76], poor accessibility [49,61,63,77], and interoperability problems [70] (8 mentions)	Flexibility [44,50,51,55,57] and data interoperability [78,90] (7 mentions)
Complexity	Data-related problems (collecting, managing, processing) [49,50,52,60,77,90], technical challenges [41,44,49,63,80,83], and overall complexity [53,80,88] (15 mentions)	None reported (0 mentions)
Design quality and usability[b]	Poor data quality [38,42,51,77,90], poor design [40,44,49,63,77,84], and difficult to use system (eg, password problems, slow speed, functionality) [42,44,46,51,56,58,78,84,90] (20 mentions)	Good data quality [56,72], good design [46], good data visualization [51,52,70,76,77], good content (eg, specific) [72], and ease of use [51,58,61] (12 mentions)
Cost	Cost of implementation [47,48] and cost of the internet [46] (3 mentions)	Technology reduces costs [67,87] (2 mentions)

[a]HCP: health care provider.

[b]The definitions have been modified from the original Consolidated Framework for Implementation Research construct codebook to match the context of this study.

The relative advantage of new technology was mentioned 54 times in the included studies, being the most frequently reported facilitator. The advantages of HIT were increased accessibility [42,45,49,75,77,85,90], 24/7 real-time access [42,49,75,77,85], and being able to acquire up-to-date information at a convenient time [90], which helped patients feel safe [45]. HIT also lessened administrative work for patients and HCPs, such as scheduling and managing appointments [42,75], organizing refill/reauthorization reminders [81], and managing data [52]. Overall, HIT was convenient [49] and helped stakeholders save time [46]. In addition, stakeholders viewed HIT as a valuable instrument for improving the quality of care [45,46,49,75,80,90]. Many examples mentioned how HIT helped improve the quality of face-to-face conversations between HCPs and patients [46,49,75]. It also helped continuous care of medical conditions [45], speed of communication [90], and prevention of medical errors [87,90]. However, HIT was sometimes noted as unhelpful [49,51,72,76] or even provoking negative emotions in the process of managing medical data [38,39]. Adequate adaptability that enables HIT to be tailored to meet various needs was revealed as a facilitator, while inapplicability [50,58,76] and poor accessibility [49,77] acted as barriers. Allowing patient choice over default settings [57], clinician autonomy and flexibility [50,51,55], and up-to-date information contributed

to adaptability [44]. "Complexity," which is the perceived difficulty that hindered the use of the system, was noted several times. Especially, data management problems, such as collecting, managing, and processing data [49,50,52,60,77,90], and frequent technical challenges [41,44,49,63,80,83] were important.

Design quality and usability was the most mentioned barrier (20 times). Inaccurate or incomplete data [38,42,51,69,77,88,90] and poor user interface or inadequate design [49,77] of app/program features were noted (eg, "prompt overload" [40] and "wordiness" [44]). Difficulties in using the system, such as frequent password problems [42,44,46], slow speed of the system [51,56,58,90], and lack of functionality, acted as barriers [51]. On the other hand, good data quality [56,72], good design [46], good data visualization [51,52,70,76,77], good content (eg, specific) [72], and good system usability [51,58] encouraged the use of HIT.

There were differing views regarding the cost of deploying HIT. Several articles regarded the expenses needed for implementing HIT as expensive and burdensome [47,48]. However, other papers suggested that using HIT could save money by lowering health care costs [67,87].

Part 2: Stakeholder Analysis

We have mapped the barriers (Table 3) and facilitators (Table 4) by the integrated framework. The references of each factor are indicated in Multimedia Appendix 9. Table 5 summarizes the numbers of times the barrier and facilitator codes in the category emerged in the selected papers.

Table 3. Stakeholder analysis with the integrated framework for barriers of health information technology implementation.

Barriers	Individual (patient)	Individual (health care professional)	Interpersonal	Organizational	Political
Outer setting					
Needs and resources	Lack of desire (n=4)[a] and lack of need (n=2)	Lack of desire (n=1) and lack of need (n=1)	N/A[b]	N/A	N/A
External policy and incentives	N/A	N/A	N/A	N/A	Regulation concerns (n=2), government policies (n=1), and lack of health system support (n=1)
Inner setting					
Structural characteristics	N/A	N/A	N/A	Organizational issues (n=4), unclear responsibilities (n=4), and organizational conflicts (n=1)	N/A
Networks and communications	N/A	N/A	Lack of connection with peers (n=1) and lack of trust (n=1)	N/A	N/A
Implementation climate	Feels like work (n=3) and competing priorities (n=2)	Competing priorities (n=3)	N/A	Tension for change (n=1), lack of fit with existing workflow (n=3), competing priorities (n=3), and lack of reimbursement (n=2)	N/A
Readiness to implementation	Lack of computer or internet (n=5), lack of financial resources (n=1), and lack of training (n=1)	Lack of time (n=7)	Lack of assistance (n=3)	Lack of leadership engagement (n=1), lack of administrative support (n=1), lack of infrastructure and equipment (n=6), lack of financial resources (n=3), lack of workforce (n=3), and increased workload (n=3)	N/A
Privacy and confidentiality	Privacy concern (n=5)	Privacy concern (n=2)	N/A	N/A	Political regulations (n=1)
Characteristics of individuals					
Knowledge and beliefs	Concerns on diminishing interaction with HCPs[c] (n=1), high expectations (n=2), lack of knowledge (n=3), and preconceived beliefs (n=3)	Lack of knowledge (n=2), past negative experience (n=2), negative attitude (n=1), resistance toward change (n=2), and concern on patient's role (n=1)	N/A	N/A	N/A
Self-efficacy	Health literacy (n=7) and lack of digital skills (n=10)	Lack of digital skills (n=2)	N/A	N/A	N/A
Other	Cognitive impairment (n=1), financial status (n=1), literacy (n=4), passive attitude (n=1), physical impairment (n=1), and inadequate knowledge of own health (n=2)	Older age (n=2) and poor communication style (n=1)	N/A	N/A	N/A
Process					
Planning	N/A	N/A	N/A	N/A	Lack of long-term plans (n=1)

Barriers	Individual (patient)	Individual (health care professional)	Interpersonal	Organizational	Political
Engaging	N/A	Lack of HCP engagement (n=2)	Lack of patient-provider engagement (n=1)	Lack of organizational commitment (n=1)	N/A
Executing	N/A	N/A	Lack of cooperation (n=1)	N/A	N/A

[a]Throughout the table, “n” refers to the number of times a code emerged in all the selected papers.

[b]N/A: not applicable.

[c]HCP: health care provider.

Table 4. Stakeholder analysis with the integrated framework for facilitators of health information technology implementation.

Facilitators	Individual (patient)	Individual (health care professional)	Interpersonal	Organizational	Political
Outer setting					
Needs and resources	Need for management and information (n=1)[a] and self-motivation (n=2)	Motivation to change (n=1)	N/A[b]	N/A	N/A
Cosmopolitanism	N/A	N/A	Positive experience of early adopters (n=2)	N/A	N/A
Peer pressure	N/A	N/A	N/A	Peer pressure (n=1)	N/A
External policy and incentives	N/A	N/A	N/A	N/A	Laws and regulations (n=1)
Inner setting					
Networks and communications	N/A	N/A	Trusted relationship (n=1) and communication (n=1)	N/A	N/A
Culture	N/A	N/A	N/A	Innovation-oriented culture (n=1)	N/A
Implementation climate	Match workflow (n=1)	N/A	N/A	Integration into workflow (n=3)	N/A
Readiness to implementation	Conducive environment (n=1) and patient education (n=2)	Training (n=3)	N/A	Administrative support (n=2), adequate infrastructure (n=2), adequate financial resources (n=1), and technical support (n=2)	N/A
Privacy and confidentiality	N/A	N/A	N/A	Adequate management of data (n=3)	N/A
Characteristics of individuals					
Knowledge and beliefs	Adequate knowledge base (n=2)	Positive attitude (n=1)	N/A	N/A	N/A
Self-efficacy	Adequate health literacy (n=1)	N/A	N/A	N/A	N/A
Other	N/A	Good communication style (n=1)	N/A	N/A	N/A
Process					
Planning	Strategic implementation process (n=1)	N/A	N/A	N/A	N/A
Engaging	HCP[c] engagement (n=1)	Physician's suggestion (n=5) and family support (n=1)	Identify and nurture champion (n=1)	N/A	N/A
Executing	N/A	Cooperation (n=3) and patient-provider communication (n=2)	Use pre-existing relationships (n=1)	N/A	N/A
Reflecting and evaluating	N/A	N/A	Feedback from provider (n=1)	Feedback (n=2) and regular monitoring (n=1)	N/A

[a]Throughout the table, "n" refers to the number of times a code emerged in all the selected papers.

[b]N/A: not applicable.

[c]HCP: health care provider.

Table 5. Summary of the stakeholder analysis with the integrated framework.

Variable	Individual (patient), n[a]		Individual (health care professional), n[a]		Interpersonal, n[a]		Organizational, n[a]		Political, n[a]	
	B[b]	F[c]	B	F	B	F	B	F	B	F
Outer setting										
Needs and resources	6	3	2	1	0	0	0	0	0	0
Cosmopolitanism	0	0	0	0	0	2	0	0	0	0
Peer pressure	0	0	0	0	0	0	0	1	0	0
Eternal policy and interventions	0	0	0	0	0	0	0	0	4	1
Inner setting										
Structural characteristics	0	0	0	0	0	2	9	0	0	0
Networks and communications	0	0	0	0	2	0	0	0	0	0
Culture	0	0	0	0	0	0	0	1	0	0
Implementation climate	5	1	3	0	0	0	9	3	0	0
Readiness to implementation	7	3	7	3	3	0	17	7	1	0
Privacy and confidentiality	5	0	2	0	0	0	0	0	0	0
Characteristics of individuals										
Knowledge and beliefs	9	2	8	1	0	0	0	0	0	0
Self-efficacy	17	1	2	0	0	0	0	0	0	0
Other	10	0	2	1	0	0	0	0	0	0
Process										
Planning	0	0	0	0	0	0	1	1	0	0
Engaging	0	0	2	1	1	6	1	1	0	0
Executing	0	0	0	0	1	6	0	1	0	0
Reflecting and evaluating	0	0	0	0	0	0	0	1	0	3

[a]The number of times the barrier/facilitator codes in the category emerged.

[b]B: barrier.

[c]F: facilitator.

Individual (Patient)-Characteristics of Individuals

The individual factors of patients were the most reported barriers and facilitators. Factors associated with the CFIR construct "characteristics of individuals," particularly self-efficacy issues, were prominent. Many patients did not have sufficient health literacy to understand the content of HIT [38,43,46,48,53,83,86] and were therefore limited in the use of HIT [46] or required assistance [83]. Lack of digital skills for using the computer and the internet challenged HIT use for both patients [41,43,46,56,57,63,66,81,83,85] and HCPs [55,87]. Some studies even revealed the existence of "computer anxiety" [43]. Sometimes patients did not have a computer or an internet connection [46,49,56]. In contrast, having adequate health literacy [46] and knowledge [38,53] acted as facilitators.

Lack of financial resources [56], cognitive impairment [43], literacy [46,61,66,83], passive attitude [76], physical impairment [43], and inadequate knowledge of own health [53] were also barriers. On the other hand, adequate knowledge of the health system and medical data [38,53], and adequate health literacy acted as facilitators [46].

Initial knowledge and beliefs on HIT were also frequently noted. Patients did not know of the existence of HIT [42,81,85] or were not aware of the tool's functions [51]. Moreover, negative preconceived attitudes toward HIT [38,75,81], such as dislike of electronic communication methods [75,81] and misconceptions about the health care system [38], hindered them from trying something new. They were also worried that a new communication method might diminish the original communication with HCPs [46].

Individual (Patient)-Needs and Resources (Outer Setting)

Patients' lack of wants or needs served as a barrier, whereas their desire for management and drive served as a facilitator. Patients were sometimes disinterested in the self-management of their disease [56,63,66,85] and preferred medical discussions based on personal clinical encounters [79], or already had an alternative method of managing their disease [56]. Medical data tracking was often conceived as effortful and time-consuming [39,56,80]. However, patients were also frequently self-motivated in incorporating HIT into their daily lives [61,65,66].

Individual/Organizational/Political-Privacy and Confidentiality (Inner Setting)

The construct "privacy and confidentiality" was added because of the unique characteristic of HIT, that is, it deals with sensitive personal information. Patients mentioned privacy concerns as a barrier to HIT implementation [38,44,46,56,72] (eg, wary about the number of people who might have access to one's medical records [72]). HCPs were also worried about the possibility of exploiting patient data [59,90]. When sufficient measures were taken to ensure the privacy of medical data, it acted as a facilitator [44,57,72]. The perception of security was increased by features like secure messaging [72], safe storage [57], and control over privacy bounds.

Processes required for privacy and security based on stakeholder needs and political regulations may operate as roadblocks to HIT adoption. Many safeguards (eg, safe login) must be taken by organizations, and such rules considerably reduce the availability of privacy-sensitive information on the portal, affecting data quality [61].

Individual (HCP)-Readiness to Implementation (Inner Setting)

There were various individual factors of HCPs that challenged the successful implementation of HIT. HCPs indicated that they have a lack of time [48,49,51,53,58,61,66,73], they did not have enough time to adjust [48,49,53,58], or the use of HIT increased consultation time and therefore depleted time resources [51,73]. HCPs often had competing priorities [58,79] in work and perceived the newly implemented HIT as noncore work activity [58].

The most frequently mentioned facilitator was training [40,50,73], and succinct and customized information was valued [50].

Individual (HCP)-Characteristics of Individuals

Individual characteristics that held up implementation were lack of knowledge [51], past negative experience [50,63], resistance toward change [50], and poor communication style [47]. Sometimes physicians preferred traditional health care messages [79] and thought that change is unneeded [51], especially because they did not believe patients could not efficiently manage their data [66].

In contrast, having a good communication style (eg, friendly and sympathetic) [47] and a good attitude toward HIT implementation acted as a facilitator [61].

Interpersonal

Many facilitators acted through interpersonal relationships. Prior experience from other HCPs provided legitimacy and had a positive influence via professional and social networks [53,59].

For patients, physician guidance [47,48,75], recommendations [85], and feedback from HCPs [80] assisted them in using HIT and made them feel supported [53]. Patients were more likely to use HIT when it was recommended by trusted physicians [72]. Patients benefited from family support as well [74]. It operated as a barrier when the need for long-term guidance by HCPs or family members was not adequately addressed [41,43,49]. The introduction of HIT was also hampered by a lack of connection with peers (patients) [41] and a lack of trust in communicating with HCPs [56,90].

In addition, cooperation between HCPs and various stakeholders was important. HCPs stated that a team approach to decision-making [53] and sharing information between providers was useful [44].

However, a lack of coordination between vendors and the hospital [57], nurses, and providers challenged the implementation process [58]. Since interpersonal factors play an important role in HIT implementation, it was recommended to leverage existing relationships to gain momentum [59].

Organizational-Inner Setting

Underlying organizational issues [55,73,78,87] and unclear responsibility of HCPs [58,65,90] created confusion. Lack of fit with existing workflow was frequently stated [51,55,79]. When new technology did not match existing practice routines or clinic schedules, the start-up period of HIT implementation was associated with an initial drop in productivity [55]. In contrast, HIT implementation matching the workflow acted as a facilitator [58,64,76]. This highlights the importance of incorporating an optimal workflow strategy [79].

The readiness of an organization to implement HIT also played a significant role. For example, lack of administrative support [77], lack of infrastructure and equipment [40,48,57,64,87,88], lack of financial resources [48,61,87], and lack of workforce [48,61,87] were noted as barriers. Conversely, administrative support [55,61], adequate infrastructure [57,58] (eg, computer resources), adequate financial resources [50], and technical support [55] were facilitators.

Political-Outer Setting

External policies at the political level had an impact as well. Stakeholders stated their concerns with the Health Insurance Portability and Accountability Act (HIPAA) regulations [52], which govern the privacy and security of personal data. There may also be some delays in the implementation of HIT that may benefit organizations owing to government policies [87]. Facilitating rules and regulations can be advantageous, as evidenced by the support for portal implementation by the Netherlands government [61]. On the other hand, deploying HIT was hampered by a lack of government and health care system support [67].

Discussion

Principal Findings

This review identified various barriers and facilitators of the implementation of HIT programs for NCD management. We conducted the analysis in 2 parts. In part 1, we focused on the inherent characteristics of HIT interventions. A relative advantage to the existing health care system was most frequently reported as a facilitator. Especially, convenience, improvement of the quality of care, and improvement in accessibility were considered useful. Design quality and usability issues, such as difficulty in using the system and data quality, were the most

prominent barriers. Tackling these practical issues would be crucial in the implementation process.

In part 2, we used the novel integrated framework to indicate the human factors of implementation. Individual factors of patients related to self-efficacy were the most noted barriers. Adequate knowledge of the health system, medical data, and adequate health literacy acted as facilitators. HCPs often indicated that they have a lack of time, while training was the most quoted facilitator. At the interpersonal level, the social relationships that support the implementation process were crucial, such as the prior experience of peers, communication with HCPs, and support from family members. At the organizational level, lack of fit with existing workflow acted as a barrier, while adequate infrastructure, technical support, and financial resources were facilitators. At the political level, regulation concerns were mentioned, but facilitating rules and regulations can help implementation.

Therefore, internal technology factors of HIT and external human factors of stakeholders are both very important to the implementation. Policymakers and relevant stakeholders should not focus on only 1 side but recognize all aspects of change to maximize the probability of success.

Comparison With Prior Work

Our findings concur with other reviews on the implementation of HIT [20,22,91,92]. Yet, previous reviews did not focus on NCD management and mostly listed the barriers and facilitators without structurization. For example, Finkelstein et al [92] mentioned 9 barriers (lack of usability, old age, education, cognitive impairment, workflow issues, etc) and 9 facilitators (perceived usefulness, efficiency, availability, etc) of HIT for patient-centered care. The importance of health literacy and being able to use the software has also been mentioned [10,93]. The interpersonal, organizational, and political factors we identified are in line with other studies that emphasized the importance of social relationships and human factors. For instance, a review on digital health interventions stated that social support affects patient engagement and recruitment [94]. However, 1 study reported that social influences have no significant effects on health care technology acceptance [95]. Further studies should try to understand the extent and pathway of social relationships in HIT implementation.

Usability has been emphasized as a critical factor in other HIT-related studies. A recent analysis financed by the Agency for Healthcare Research and Quality found significant flaws in the procedures, methods, and application of standards and best practices in the areas of usability and human aspects among certified EHR vendors [96]. EHRs must be used efficiently and effectively as they increasingly become a major tool for patient care. Moreover, usability difficulties for HIT in NCD management are consistent with existing usability research. One of the most used usability evaluation tools in information technology is the Health Information Technology Usability Evaluation Scale (Health-ITUES) [97]. Although the original Health-ITUES focused on mHealth technology, several aspects of our analysis overlap. “Improving the quality of life,” “having positive influence,” and “perceived usefulness” were mentioned as relative advantages for HIT in our study. Concepts related to the category “perceived ease of use” and “user control” were coded to the CFIR construct “design quality and usability.” This resemblance emphasizes the importance of usability difficulties in the acceptance of new technologies.

The individual barriers identified in this review are consistent with the analysis of Sun et al regarding what can aggravate the digital divide (limited technical infrastructure, lack of digital literacy, financial resources, and lack of access to digital hardware) [98]. The UN Secretary-General’s high-level panel on digital cooperation has also warned of rapid digitization leaving marginalized people behind [99]. The shortage of digital infrastructure in developing countries makes it vital to put the digital divide in context when developing HIT-related health policies, considering that only 45% of people are connected to the internet in developing countries [100]. The age-related digital divide is also an emerging problem. As our review and other reports have shown [43,46,86,92], many older patients fear technology and need detailed guidance. Policymakers should not neglect these issues of inequality and should pay attention to the underlying socioeconomic conditions in every step of the planning and implementation of HIT.

We have included the construct “privacy and confidentiality” within the “inner setting” of our integrated framework. The issue of privacy is a heated discussion in studies on information technology. The problem of dealing with personal health information has been identified in many countries, and the current legal framework is sometimes hard to match with the system [101]. This has affected the new legislation, for example, the HIPAA in the United States in 2013 and the General Data Protection Act in the European Union in 2016. Organizations could discuss deidentification methods of health information such as anonymization and pseudonymization. The acquisition of consent is also a complicated issue. For example, the usage of data should be differentiated depending on whether patients agreed to give their medical information for only treatment or for both research and treatment purposes. For now, “opt-in” (users taking affirmative action to offer their consent) is standard. “Opt-out” choices from national data (users taking action to withdraw consent) have also been offered in the United Kingdom [102], and this could also be considered in future HIT implementations.

Five research gaps have been identified through this review. First, most studies only mentioned patients and physicians. Other stakeholders, such as vendors, service providers, government officials, and administrative workforces, should be addressed in future research. Second, a great majority of HIT interventions targeted the diabetes population. This may be expected since diabetes involves the strictest self-management, such as weekly blood glucose testing. Nonetheless, there is an evident lack of research on the management of other chronic diseases, such as obesity and mental diseases. Further research in this area is warranted. Third, little evidence exists on the challenges of the long-term implementation of programs. Most studies included in this review covered implementations that were followed up for a short term. Fourth, the included studies might have been biased in the selection of study participants because they rarely used random sampling. More rigorous methods should be used, and response rates and reasons for

unavailability or decline of participation should be reported. In addition, as our prior discussion on the digital divide implies, participants who have access to ongoing HIT programs might be inclined to have a higher socioeconomic status. Therefore, further studies should consider how to sufficiently represent older, socioeconomically disadvantaged, and other underrepresented groups. The final gap results from the underrepresentation of various countries, which may limit the generalizability of our findings. Most studies were conducted in the United States and other high-income countries. Extensive research on the implementation strategies of HIT in LMICs is necessary.

Strengths and Limitations

This review has several strengths. First, to the best of our knowledge, this review is the first systematic review on the topic of HIT for NCD management. Second, our search strategy included as many eligible studies as possible, and double screening was performed at all stages. Third, we developed the integrated framework based on 2 widely recognized frameworks [26,37], which are comprehensive and detailed. Fourth, the quality of studies was assessed, but we did not restrict the inclusion of studies based on quality in order to capture as much literature as possible.

There were some limitations of this study. First, although the quality of the included studies was generally good, some studies were of low quality. The low-quality studies were not used to draw conclusions and had little effect on our overall findings. Second, since the included studies were about different types of HIT interventions and stakeholders, there could be limitations in applying the results to a specific setting. Finally, the perceived importance of facilitators and barriers in this study may not always correspond with the actual importance, and some factors may be more hypothetical. The reported factors may also have been influenced by publication bias.

Conclusions

Internal factors of HIT and external human factors of implementation interplay in the implementation of HIT for chronic disease management. Among the characteristics of the intervention, having a relative advantage over existing health care was the most noted facilitator, while poor usability was the most reported barrier. In our stakeholder analysis undertaken by the integrated framework, health literacy and lack of digital skills were identified as key barriers. Various interpersonal and organizational factors were crucial (eg, physicians' suggestions, cooperation, adequate management of data, and addressing privacy concerns). Implementation strategies of HIT could be improved by studying these barriers and facilitators. Further research should focus on studying various stakeholders, such as service providers and administrative workforces; various disease populations, such as those with obesity and mental diseases; and various countries, including LMICs.

Authors' Contributions

MS was responsible for the study design, screening, quality appraisals, data extraction, synthesis of results, and writing of the manuscript. JH provided input for the development of the eligibility criteria, screened titles and abstracts, screened full texts, conducted data extraction, and contributed and provided comments on several drafts of the manuscript. JJ, QZ, and YC advised and revised several drafts of the manuscript. HC provided helpful comments on several drafts of the manuscript. ZL advised on the design of the study, eligibility criteria, and search methods, and provided comments and feedback on several drafts of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA checklist.
[DOCX File , 16 KB - jmir_v24i7e37338_app1.docx]

Multimedia Appendix 2
Search syntax.
[DOCX File , 24 KB - jmir_v24i7e37338_app2.docx]

Multimedia Appendix 3
Data extraction form.
[DOCX File , 19 KB - jmir_v24i7e37338_app3.docx]

Multimedia Appendix 4
Quality appraisal domains by study methodology.
[DOCX File , 14 KB - jmir_v24i7e37338_app4.docx]

Multimedia Appendix 5

Critical appraisal of studies.
[DOCX File , 29 KB - jmir_v24i7e37338_app5.docx]

Multimedia Appendix 6
Enhancing Transparency in Reporting the Synthesis of Qualitative Research checklist.
[DOCX File , 17 KB - jmir_v24i7e37338_app6.docx]

Multimedia Appendix 7
Study characteristics.
[DOCX File , 53 KB - jmir_v24i7e37338_app7.docx]

Multimedia Appendix 8
Descriptive theme definitions and representative quotes.
[DOCX File , 47 KB - jmir_v24i7e37338_app8.docx]

Multimedia Appendix 9
Barriers and facilitators of the implementation of health information technology by the integrated framework.
[DOCX File , 189 KB - jmir_v24i7e37338_app9.docx]

References

1. Noncommunicable diseases. Pan American Health Organization. URL: https://www.paho.org/en/topics/noncommunicable-diseases [accessed 2021-11-15]
2. Global Health Estimates 2020: Disease burden by Cause, Age, Sex, by Country and by Region, 2000-2019. World Health Organization. 2020. URL: https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/global-health-estimates-leading-causes-of-dalys [accessed 2022-06-30]
3. Noncommunicable diseases. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2021-11-17]
4. Marchibroda J. The impact of health information technology on collaborative chronic care management. J Manag Care Pharm 2008 Mar;14(2 Suppl):S3-11. [Medline: 18331114]
5. Health IT: Advancing America's Health Care. The Office of the National Coordinator for Health Information Technology. URL: https://www.healthit.gov/sites/default/files/pdf/health-information-technology-fact-sheet.pdf [accessed 2021-11-15]
6. Health IT for Improved Chronic Disease Management. Agency for Healthcare Research and Quality. URL: https://digital.ahrq.gov/sites/default/files/docs/page/08-0084_cdm_0.pdf [accessed 2021-11-17]
7. Samoocha D, Bruinvels DJ, Elbers NA, Anema JR, van der Beek AJ. Effectiveness of web-based interventions on patient empowerment: a systematic review and meta-analysis. J Med Internet Res 2010 Jun 24;12(2):e23 [FREE Full text] [doi: 10.2196/jmir.1286] [Medline: 20581001]
8. Webb R, Uddin N, Ford E, Easter A, Shakespeare J, Roberts N, et al. Barriers and facilitators to implementing perinatal mental health care in health and social care settings: a systematic review. The Lancet Psychiatry 2021 Jun;8(6):521-534. [doi: 10.1016/s2215-0366(20)30467-3]
9. Parasrampuria Sonal, Henry Jawanna. Hospitals' Use of Electronic Health Records Data, 2015-2017. ONC Data Brief No. 46.: The Office of the National Coordinator for Health Information Technology URL: https://www.healthit.gov/sites/default/files/page/2019-04/AHAEHRUseDataBrief.pdf [accessed 2022-06-30]
10. Anwar F, Shamim A. Barriers in adoption of health information technology in developing societies. IJACSA 2011;2(8). [doi: 10.14569/ijacsa.2011.020808]
11. Biruk S, Yilma T, Andualem M, Tilahun B. Health Professionals' readiness to implement electronic medical record system at three hospitals in Ethiopia: a cross sectional study. BMC Med Inform Decis Mak 2014 Dec 12;14(1):115 [FREE Full text] [doi: 10.1186/s12911-014-0115-5] [Medline: 25495757]
12. Willyard C. Focus on electronic health records. Electronic records pose dilemma in developing countries. Nat Med 2010 Mar;16(3):249. [doi: 10.1038/nm0310-249a] [Medline: 20208497]
13. World Health Organization. Electronic health records : manual for developing countries. WHO Regional Office for the Western Pacific. 2006. URL: https://apps.who.int/iris/handle/10665/207504 [accessed 2022-06-30]
14. Khoja S, Scott RE, Casebeer AL, Mohsin M, Ishaq A, Gilani S. e-Health readiness assessment tools for healthcare institutions in developing countries. Telemed J E Health 2007 Aug;13(4):425-431. [doi: 10.1089/tmj.2006.0064] [Medline: 17848110]
15. Ajami S, Ketabi S, Isfahani S, Heidari A. Readiness assessment of electronic health records implementation. Acta Inform Med 2011 Dec;19(4):224-227 [FREE Full text] [doi: 10.5455/aim.2011.19.224-227] [Medline: 23407861]
16. Feller DJ, Lor M, Zucker J, Yin MT, Olender S, Ferris DC, et al. An investigation of the information technology needs associated with delivering chronic disease care to large clinical populations. Int J Med Inform 2020 May;137:104099. [doi: 10.1016/j.ijmedinf.2020.104099] [Medline: 32088558]

17. Marquard J. Human factors and organizational issues in health informatics: Innovations and opportunities. Yearb Med Inform 2021 Aug 03;30(1):91-99 [FREE Full text] [doi: 10.1055/s-0041-1726511] [Medline: 34479382]
18. Carayon P, Hoonakker P. Human factors and usability for health information technology: Old and new challenges. Yearb Med Inform 2019 Aug 16;28(1):71-77 [FREE Full text] [doi: 10.1055/s-0039-1677907] [Medline: 31419818]
19. Abbott PA, Weinger MB. Health information technology:Fallacies and Sober realities - Redux A homage to Bentzi Karsh and Robert Wears. Appl Ergon 2020 Jan;82:102973. [doi: 10.1016/j.apergo.2019.102973] [Medline: 31677422]
20. Amante DJ, Hogan TP, Pagoto SL, English TM. A systematic review of electronic portal usage among patients with diabetes. Diabetes Technol Ther 2014 Nov;16(11):784-793. [doi: 10.1089/dia.2014.0078] [Medline: 24999599]
21. Kaufman N. Using health information technology to prevent and treat diabetes. Int J Clin Pract Suppl 2012 Mar;66(175):40-48. [doi: 10.1111/j.1742-1241.2011.02853.x] [Medline: 22308988]
22. Osborn CY, Mayberry LS, Mulvaney SA, Hess R. Patient web portals to improve diabetes outcomes: a systematic review. Curr Diab Rep 2010 Dec 2;10(6):422-435 [FREE Full text] [doi: 10.1007/s11892-010-0151-1] [Medline: 20890688]
23. Sheon AR, Bolen SD, Callahan B, Shick S, Perzynski AT. Addressing disparities in diabetes management through novel approaches to encourage technology adoption and use. JMIR Diabetes 2017 Jul 13;2(2):e16 [FREE Full text] [doi: 10.2196/diabetes.6751] [Medline: 30291090]
24. Sun R, Korytkowski MT, Sereika SM, Saul MI, Li D, Burke LE. Patient portal use in diabetes management: Literature review. JMIR Diabetes 2018 Nov 06;3(4):e11199 [FREE Full text] [doi: 10.2196/11199] [Medline: 30401665]
25. Dehnavi Z, Ayatollahi H, Hemmat M, Abbasi R. Health information technology and diabetes management: A review of motivational and inhibitory factors. Curr Diabetes Rev 2021 Mar 12;17(3):268-279. [doi: 10.2174/1573399816666200719012849] [Medline: 32682380]
26. 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;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
27. Esponda GM, Hartman S, Qureshi O, Sadler E, Cohen A, Kakuma R. Barriers and facilitators of mental health programmes in primary care in low-income and middle-income countries. The Lancet Psychiatry 2020 Jan;7(1):78-92. [doi: 10.1016/s2215-0366(19)30125-7]
28. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, 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]
29. Cochrane Handbook for Systematic Reviews of Interventions. 2021. URL: https://training.cochrane.org/handbook/archive/v6.2 [accessed 2022-06-30]
30. Joanna Briggs Institute. Critical Appraisal Checklist for Qualitative Research. JBI. URL: https://jbi.global/sites/default/files/2021-10/Checklist_for_Qualitative_Research.docx [accessed 2022-06-30]
31. Hong Q, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. he Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. Education for Information. 2018 Nov 12. URL: http://mixedmethodsappraisaltoolpublic.pbworks.com/w/page/24607821/FrontPage [accessed 2021-11-15]
32. Critical Appraisal Checklist for Cross-Sectional Study. Center for Evidence Based Management. 2014. URL: https://cebma.org/wp-content/uploads/Critical-Appraisal-Questions-for-a-Cross-Sectional-Study-July-2014-1.pdf [accessed 2021-11-15]
33. Tong A, Flemming K, McInnes E, Oliver S, Craig J. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. BMC Med Res Methodol 2012 Nov 27;12(1):181 [FREE Full text] [doi: 10.1186/1471-2288-12-181] [Medline: 23185978]
34. Carroll C, Booth A, Leaviss J, Rick J. "Best fit" framework synthesis: refining the method. BMC Med Res Methodol 2013 Mar 13;13(1):37 [FREE Full text] [doi: 10.1186/1471-2288-13-37] [Medline: 23497061]
35. Glasgow RE, Harden SM, Gaglio B, Rabin B, Smith ML, Porter GC, et al. RE-AIM planning and evaluation framework: Adapting to new science and practice with a 20-year review. Front Public Health 2019;7:64 [FREE Full text] [doi: 10.3389/fpubh.2019.00064] [Medline: 30984733]
36. 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]
37. Ferlie EB, Shortell SM. Improving the quality of health care in the United Kingdom and the United States: a framework for change. Milbank Q 2001 May;79(2):281-315 [FREE Full text] [doi: 10.1111/1468-0009.00206] [Medline: 11439467]
38. Ancker JS, Witteman HO, Hafeez B, Provencher T, Van de Graaf M, Wei E. The invisible work of personal health information management among people with multiple chronic conditions: qualitative interview study among patients and providers. J Med Internet Res 2015 Jun 04;17(6):e137 [FREE Full text] [doi: 10.2196/jmir.4381] [Medline: 26043709]
39. Ancker JS, Witteman HO, Hafeez B, Provencher T, Van de Graaf M, Wei E. "You Get Reminded You're a Sick Person": Personal data tracking and patients with multiple chronic conditions. J Med Internet Res 2015 Aug 19;17(8):e202 [FREE Full text] [doi: 10.2196/jmir.4209] [Medline: 26290186]
40. Dikomitis L, Green T, Macleod U. Embedding electronic decision-support tools for suspected cancer in primary care: a qualitative study of GPs' experiences. Prim Health Care Res Dev 2015 Mar 3;16(06):548-555. [doi: 10.1017/s1463423615000109]

41. Fontil V, McDermott K, Tieu L, Rios C, Gibson E, Sweet CC, et al. Adaptation and feasibility study of a digital health program to prevent diabetes among low-income patients: Results from a partnership between a digital health company and an academic research team. J Diabetes Res 2016;2016:8472391 [FREE Full text] [doi: 10.1155/2016/8472391] [Medline: 27868070]
42. Hess R, Bryce CL, Paone S, Fischer G, McTigue KM, Olshansky E, et al. Exploring challenges and potentials of personal health records in diabetes self-management: implementation and initial assessment. Telemed J E Health 2007 Oct;13(5):509-517. [doi: 10.1089/tmj.2006.0089] [Medline: 17999613]
43. Lober WB, Zierler B, Herbaugh A, Shinstrom SE, Stolyar A, Kim EH, et al. Barriers to the use of a personal health record by an elderly population. AMIA Annu Symp Proc 2006:514-518 [FREE Full text] [Medline: 17238394]
44. Marchak JG, Cherven B, Williamson Lewis R, Edwards P, Meacham LR, Palgon M, et al. User-centered design and enhancement of an electronic personal health record to support survivors of pediatric cancers. Support Care Cancer 2020 Aug 18;28(8):3905-3914 [FREE Full text] [doi: 10.1007/s00520-019-05199-w] [Medline: 31853699]
45. Ralston JD, Revere D, Robins LS, Goldberg HI. Patients' experience with a diabetes support programme based on an interactive electronic medical record: qualitative study. BMJ 2004 May 15;328(7449):1159 [FREE Full text] [doi: 10.1136/bmj.328.7449.1159] [Medline: 15142919]
46. 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]
47. Tong WT, Ng CJ, Lee YK, Lee PY. What will make patients use a patient decision aid? A qualitative study on patients' perspectives on implementation barriers and facilitators. J Eval Clin Pract 2020 Jun 21;26(3):755-764. [doi: 10.1111/jep.13161] [Medline: 31115132]
48. Tong WT, Lee YK, Ng CJ, Lee PY. Factors influencing implementation of an insulin patient decision aid at public health clinics in Malaysia: A qualitative study. PLoS One 2020 Dec 30;15(12):e0244645 [FREE Full text] [doi: 10.1371/journal.pone.0244645] [Medline: 33378349]
49. Urowitz S, Wiljer D, Dupak K, Kuehner Z, Leonard K, Lovrics E, et al. Improving diabetes management with a patient portal: a qualitative study of diabetes self-management portal. J Med Internet Res 2012 Nov 30;14(6):e158 [FREE Full text] [doi: 10.2196/jmir.2265] [Medline: 23195925]
50. Varonen H, Kortteisto T, Kaila M, EBMeDS Study Group. What may help or hinder the implementation of computerized decision support systems (CDSSs): a focus group study with physicians. Fam Pract 2008 Jun 25;25(3):162-167. [doi: 10.1093/fampra/cmn020] [Medline: 18504253]
51. Wan Q, Makeham M, Zwar NA, Petche S. Qualitative evaluation of a diabetes electronic decision support tool: views of users. BMC Med Inform Decis Mak 2012 Jul 03;12(1):61 [FREE Full text] [doi: 10.1186/1472-6947-12-61] [Medline: 22759239]
52. Wang J, Siminerio LM. Educators' insights in using chronicle diabetes: a data management system for diabetes education. Diabetes Educ 2013 Feb 20;39(2):248-254. [doi: 10.1177/0145721713475844] [Medline: 23427239]
53. Yu CH, Ke C, Jovicic A, Hall S, Straus SE, IP-SDM Team. Beyond pros and cons - developing a patient decision aid to cultivate dialog to build relationships: insights from a qualitative study and decision aid development. BMC Med Inform Decis Mak 2019 Sep 18;19(1):186 [FREE Full text] [doi: 10.1186/s12911-019-0898-5] [Medline: 31533828]
54. Zwaanswijk M, Ploem M, Wiesman F, Verheij R, Friele R, Gevers J. Understanding health care providers' reluctance to adopt a national electronic patient record: an empirical and legal analysis. Med Law 2013 Mar;32(1):13-31. [Medline: 23781762]
55. Trivedi MH, Daly EJ, Kern JK, Grannemann BD, Sunderajan P, Claassen CA. Barriers to implementation of a computerized decision support system for depression: an observational report on lessons learned in "real world" clinical settings. BMC Med Inform Decis Mak 2009 Jan 21;9(1):6 [FREE Full text] [doi: 10.1186/1472-6947-9-6] [Medline: 19159458]
56. Fuji KT, Abbott AA, Galt KA. A qualitative study of how patients with type 2 diabetes use an electronic stand-alone personal health record. Telemed J E Health 2015 Apr;21(4):296-300 [FREE Full text] [doi: 10.1089/tmj.2014.0084] [Medline: 25614996]
57. Marquard JL, Garber L, Saver B, Amster B, Kelleher M, Preusse P. Overcoming challenges integrating patient-generated data into the clinical EHR: lessons from the CONtrolling Disease Using Inexpensive IT--Hypertension in Diabetes (CONDUIT-HID) Project. Int J Med Inform 2013 Oct;82(10):903-910. [doi: 10.1016/j.ijmedinf.2013.04.009] [Medline: 23800678]
58. Saleem JJ, Patterson ES, Militello L, Render ML, Orshansky G, Asch SM. Exploring barriers and facilitators to the use of computerized clinical reminders. J Am Med Inform Assoc 2005 Jul 01;12(4):438-447. [doi: 10.1197/jamia.m1777]
59. Heider AR, Maloney N, Satchidanand N, Allen G, Mueller R, Gangloff S, et al. Developing a communitywide electronic health record disease registry in primary care practices: lessons learned from the Western new york beacon community. EGEMS (Wash DC) 2014 Sep 04;2(3):1089 [FREE Full text] [doi: 10.13063/2327-9214.1089] [Medline: 25848616]
60. Solberg LI, Ferguson R, Ohnsorg KA, Crain AL, Williams MD, Ziegenfuss JY, et al. The challenges of collecting and using patient care data from diverse care systems: Lessons From COMPASS. Am J Med Qual 2017 Oct 20;32(5):494-499. [doi: 10.1177/1062860616674272] [Medline: 28862026]

61. Kooij L, Groen WG, van Harten WH. Barriers and facilitators affecting patient portal implementation from an organizational perspective: Qualitative study. J Med Internet Res 2018 May 11;20(5):e183 [FREE Full text] [doi: 10.2196/jmir.8989] [Medline: 29752253]
62. McBride D, Dohan D, Handley MA, Powe NR, Tuot DS. Developing a CKD registry in primary care: provider attitudes and input. Am J Kidney Dis 2014 Apr;63(4):577-583 [FREE Full text] [doi: 10.1053/j.ajkd.2013.10.012] [Medline: 24295612]
63. 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]
64. Pratt R, Saman DM, Allen C, Crabtree B, Ohnsorg K, Sperl-Hillen JM, et al. Assessing the implementation of a clinical decision support tool in primary care for diabetes prevention: a qualitative interview study using the Consolidated Framework for Implementation Science. BMC Med Inform Decis Mak 2022 Jan 15;22(1):15 [FREE Full text] [doi: 10.1186/s12911-021-01745-x] [Medline: 35033029]
65. Abidi S, Vallis M, Piccinini-Vallis H, Imran SA, Abidi SSR. Diabetes-related behavior change knowledge transfer to primary care practitioners and patients: Implementation and evaluation of a digital health platform. JMIR Med Inform 2018 Apr 18;6(2):e25 [FREE Full text] [doi: 10.2196/medinform.9629] [Medline: 29669705]
66. Baudendistel I, Winkler E, Kamradt M, Brophy S, Längst G, Eckrich F, et al. The patients' active role in managing a personal electronic health record: a qualitative analysis. Support Care Cancer 2015 Sep;23(9):2613-2621. [doi: 10.1007/s00520-015-2620-1] [Medline: 25652149]
67. Dehnavi Z, Ayatollahi H, Hemmat M, Abbasi R. Patients' and phsysicians' perspectives about using health information technology in diabetes management in Iran: A qualitative study. Perspect Health Inf Manag 2021;18(4):1i [FREE Full text] [Medline: 34975358]
68. Desai AV, Agarwal R, Epstein AS, Kuperman GJ, Michael CL, Mittelstaedt H, et al. Needs and perspectives of cancer center stakeholders for access to patient values in the electronic health record. JCO Oncology Practice 2021 Oct;17(10):e1524-e1536. [doi: 10.1200/op.20.00644]
69. Groenhof TKJ, Rittersma ZH, Bots ML, Brandjes M, Jacobs JJL, Grobbee DE, Members of the UCC-CVRM Study Group. A computerised decision support system for cardiovascular risk management 'live' in the electronic health record environment: development, validation and implementation-the Utrecht Cardiovascular Cohort Initiative. Neth Heart J 2019 Sep 1;27(9):435-442 [FREE Full text] [doi: 10.1007/s12471-019-01308-w] [Medline: 31372838]
70. Hans PK, Gray CS, Gill A, Tiessen J. The provider perspective: investigating the effect of the Electronic Patient-Reported Outcome (ePRO) mobile application and portal on primary care provider workflow. Prim Health Care Res Dev 2017 Sep 13;19(02):151-164. [doi: 10.1017/s1463423617000573]
71. Janssen A, Donnelly C, Elder E, Pathmanathan N, Shaw T. Electronic medical record implementation in tertiary care: factors influencing adoption of an electronic medical record in a cancer centre. BMC Health Serv Res 2021 Jan 06;21(1):23 [FREE Full text] [doi: 10.1186/s12913-020-06015-6] [Medline: 33407449]
72. Brown S, Grijalva C, Ferrara A. Leveraging EHRs for patient engagement: perspectives on tailored program outreach. Am J Manag Care 2017 Jul 01;23(7):e223-e230 [FREE Full text] [Medline: 28850790]
73. Marcolino MS, Oliveira JAQ, Cimini CCR, Maia JX, Pinto VSOA, Sá TQV, et al. Development and implementation of a decision support system to improve control of hypertension and diabetes in a resource-constrained area in Brazil: Mixed methods study. J Med Internet Res 2021 Jan 11;23(1):e18872 [FREE Full text] [doi: 10.2196/18872] [Medline: 33427686]
74. Mayberry LS, Kripalani S, Rothman RL, Osborn CY. Bridging the digital divide in diabetes: family support and implications for health literacy. Diabetes Technol Ther 2011 Oct;13(10):1005-1012 [FREE Full text] [doi: 10.1089/dia.2011.0055] [Medline: 21718098]
75. 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]
76. Wildeboer A, du Pon E, Schuling J, Haaijer-Ruskamp FM, Denig P. Views of general practice staff about the use of a patient-oriented treatment decision aid in shared decision making for patients with type 2 diabetes: A mixed-methods study. Health Expect 2018 Feb 21;21(1):64-74 [FREE Full text] [doi: 10.1111/hex.12586] [Medline: 28636186]
77. Conway NT, Allardice B, Wake DJ, Cunningham SG. User experiences of an electronic personal health record for diabetes. J Diabetes Sci Technol 2019 Jul 26;13(4):744-750 [FREE Full text] [doi: 10.1177/1932296818818837] [Medline: 30585085]
78. Dixon BE, Simonaitis L, Goldberg HS, Paterno MD, Schaeffer M, Hongsermeier T, et al. A pilot study of distributed knowledge management and clinical decision support in the cloud. Artif Intell Med 2013 Sep;59(1):45-53. [doi: 10.1016/j.artmed.2013.03.004] [Medline: 23545327]
79. Grant RW, Ashburner JM, Jernigan MC, Chang J, Borowsky LH, Chang Y, et al. Randomized trial of a health IT tool to support between-visit-based laboratory monitoring for chronic disease medication prescriptions. J Gen Intern Med 2015 May 6;30(5):619-625 [FREE Full text] [doi: 10.1007/s11606-014-3152-y] [Medline: 25560319]

80. Jethwani K, Ling E, Mohammed M, Myint-U K, Pelletier A, Kvedar JC. Diabetes connect: an evaluation of patient adoption and engagement in a web-based remote glucose monitoring program. J Diabetes Sci Technol 2012 Nov 01;6(6):1328-1336 [FREE Full text] [doi: 10.1177/193229681200600611] [Medline: 23294777]
81. 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]
82. Pemu P, Josiah Willock R, Alema-Mensah E, Rollins L, Brown M, Saint Clair B, et al. Achieving health equity with e-Healthystrides©: Patient perspectives of a consumer health information technology application. Ethn Dis 2019 Jun 13;29(Supp2):393-404. [doi: 10.18865/ed.29.s2.393]
83. Tieu L, Schillinger D, Sarkar U, Hoskote M, Hahn KJ, Ratanawongsa N, et al. Online patient websites for electronic health record access among vulnerable populations: portals to nowhere? J Am Med Inform Assoc 2017 Apr 01;24(e1):e47-e54 [FREE Full text] [doi: 10.1093/jamia/ocw098] [Medline: 27402138]
84. Diaz-Garelli F, Strowd R, Ahmed T, Lycan TW, Daley S, Wells BJ, et al. What oncologists want: Identifying challenges and preferences on diagnosis data entry to reduce EHR-induced burden and improve clinical data quality. JCO Clinical Cancer Informatics 2021 Dec(5):527-540. [doi: 10.1200/cci.20.00174]
85. Ronda MCM, Dijkhorst-Oei L, Rutten GEHM. Reasons and barriers for using a patient portal: survey among patients with diabetes mellitus. J Med Internet Res 2014 Nov 25;16(11):e263 [FREE Full text] [doi: 10.2196/jmir.3457] [Medline: 25424228]
86. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, Lopez A, et al. The literacy divide: health literacy and the use of an internet-based patient portal in an integrated health system-results from the diabetes study of northern California (DISTANCE). J Health Commun 2010;15 Suppl 2:183-196 [FREE Full text] [doi: 10.1080/10810730.2010.499988] [Medline: 20845203]
87. Kabukye JK, de Keizer N, Cornet R. Assessment of organizational readiness to implement an electronic health record system in a low-resource settings cancer hospital: A cross-sectional survey. PLoS One 2020 Jun 16;15(6):e0234711 [FREE Full text] [doi: 10.1371/journal.pone.0234711] [Medline: 32544214]
88. Allain TJ, Mang'anda G, Kasiya M, Khomani P, Banda NP, Gonani A, et al. Use of an electronic medical record to monitor efficacy of diabetes care in out-patients in a central hospital in Malawi: Patterns of glycaemic control and lessons learned. Malawi Med J 2017 Dec 12;29(4):322-326 [FREE Full text] [doi: 10.4314/mmj.v29i4.8] [Medline: 29963288]
89. Desai JP, Wu P, Nichols GA, Lieu TA, O'Connor PJ. Diabetes and asthma case identification, validation, and representativeness when using electronic health data to construct registries for comparative effectiveness and epidemiologic research. Med Care 2012 Jul;50 Suppl:S30-S35 [FREE Full text] [doi: 10.1097/MLR.0b013e318259c011] [Medline: 22692256]
90. Zwaanswijk M, Verheij RA, Wiesman FJ, Friele RD. Benefits and problems of electronic information exchange as perceived by health care professionals: an interview study. BMC Health Serv Res 2011 Oct 07;11(1):256 [FREE Full text] [doi: 10.1186/1472-6963-11-256] [Medline: 21982395]
91. Adaji A, Schattner P, Jones K. The use of information technology to enhance diabetes management in primary care: a literature review. Inform Prim Care 2008 Nov 01;16(3):229-237 [FREE Full text] [doi: 10.14236/jhi.v16i3.698] [Medline: 19094410]
92. Finkelstein J, Knight A, Marinopoulos S, Gibbons MC, Berger Z, Aboumatar H, et al. Enabling patient-centered care through health information technology. Evid Rep Technol Assess (Full Rep) 2012 Jun(206):1-1531. [Medline: 24422882]
93. Omboni S, Ferrari R. The role of telemedicine in hypertension management: focus on blood pressure telemonitoring. Curr Hypertens Rep 2015 Apr 20;17(4):535. [doi: 10.1007/s11906-015-0535-3] [Medline: 25790799]
94. 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]
95. Phichitchaisopa N, Naenna T. Factors affecting the adoption of healthcare information technology. EXCLI J 2013;12:413-436 [FREE Full text] [Medline: 26417235]
96. Gibbons MC, Lowry SZ, Patterson ES. Applying human factors principles to mitigate usability issues related to embedded ssumptions in health information technology design. JMIR Hum Factors 2014 Dec 18;1(1):e3 [FREE Full text] [doi: 10.2196/humanfactors.3524] [Medline: 27025349]
97. Schnall R, Cho H, Liu J. Health information technology usability evaluation scale (Health-ITUES) for usability assessment of mobile health technology: Validation study. JMIR Mhealth Uhealth 2018 Jan 05;6(1):e4 [FREE Full text] [doi: 10.2196/mhealth.8851] [Medline: 29305343]
98. Sun N, Esom K, Dhaliwal M, Amon J. Human rights and digital health technologies. Health Hum Rights 2020 Dec;22(2):21-32 [FREE Full text] [Medline: 33390689]
99. The Age of Digital Interdependence. UN Secretary-General's High-level Panel on Digital Cooperation. URL: https://www.un.org/en/pdfs/HLP%20on%20Digital%20Cooperation%20Report%20Executive%20Summary%20-%20ENG.pdf [accessed 2021-12-20]

100. Makri A. Bridging the digital divide in health care. The Lancet Digital Health 2019 Sep;1(5):e204-e205. [doi: 10.1016/s2589-7500(19)30111-6]
101. Hasselbalch G. Making sense of data ethics. The powers behind the data ethics debate in European policymaking. Internet Policy Review 2019;8(2) [FREE Full text] [doi: 10.14763/2019.2.1401]
102. Consent. Information Commissioner's Office. URL: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/consent/ [accessed 2021-11-17]

Abbreviations

CFIR: Consolidated Framework for Implementation Research
EHR: electronic health record
HCP: health care provider
Health-ITUES: Health Information Technology Usability Evaluation Scale
HIPAA: Health Insurance Portability and Accountability Act
HIT: health information technology
LMIC: low- and middle-income country
NCD: noncommunicable disease
WHO: World Health Organization

Edited by T Leung; submitted 16.02.22; peer-reviewed by E Khoong, S Lundell; comments to author 03.05.22; revised version received 25.05.22; accepted 27.06.22; published 20.07.22.

Please cite as:
Sung M, He J, Zhou Q, Chen Y, Ji JS, Chen H, Li Z
Using an Integrated Framework to Investigate the Facilitators and Barriers of Health Information Technology Implementation in Noncommunicable Disease Management: Systematic Review
J Med Internet Res 2022;24(7):e37338
URL: https://www.jmir.org/2022/7/e37338
doi:10.2196/37338
PMID:35857364

Review

eHealth and Web-Based Interventions for Informal Carers of People With Dementia in the Community: Umbrella Review

Bethan Naunton Morgan[1], BSc, MSc; Gill Windle[2], BSc, MSc, PhD; Rebecca Sharp[3], BSc, MSc, PhD; Carolien Lamers[4], Drs, DClinPsy

[1]School of Human and Behavioural Sciences, Bangor University, Bangor, United Kingdom

[2]School of Medical and Health Sciences, Bangor University, Bangor, United Kingdom

[3]The University of Auckland, Auckland, New Zealand

[4]North Wales Clinical Psychology Programme, Bangor University, Bangor, United Kingdom

Corresponding Author:
Bethan Naunton Morgan, BSc, MSc
School of Human and Behavioural Sciences
Bangor University
Brigantia Building
Penrallt Road
Bangor
United Kingdom
Phone: 44 01248 38 8107
Email: psu2fc@bangor.ac.uk

Abstract

Background: The prevalence of dementia is increasing, and there are many associated problems that family members face as informal carers, including emotional, physical, and financial difficulties. There are benefits for a person with dementia to live at home for as long as possible, and therefore, supporting their informal carers is crucial. The growing interest in supporting carers through internet-based interventions is evidenced by the volume of systematic reviews on this topic. It is now appropriate to systematically examine this body of work and provide an overview of the literature.

Objective: This umbrella review aimed to identify the most effective internet-based intervention content and delivery method to support those caring for someone with dementia living in the community.

Methods: PsycINFO, Web of Science, CINAHL, MEDLINE, Cochrane Library, and PubMed were searched for systematic reviews examining the effectiveness of web-based interventions for informal carers of people with dementia. A total of 3 reviewers extracted data and evaluated the quality of the papers. To ascertain the extent to which the systematic reviews reported on the same evidence, the proportion of overlap between their included studies was calculated. Qualitative research findings were extracted and reported.

Results: A total of 21 papers were included in the study. The quality of the review papers was mainly rated as low to moderate, and 10% (2/21) of papers were of high quality. The findings suggest that multicomponent interventions were the most effective in supporting carers. These included combinations of cognitive behavioral therapy and relaxation strategies, educational resources, and online support groups. Interventions that were delivered on the web but included sessions with a personal element, such as telephone contact, showed the best results. When comparing the studies reviewed in all the review papers, a moderate overlap was noted. However, when comparing individual reviews with each other, they showed a high overlap of the included studies.

Conclusions: Mixed delivery methods and intervention content showed the most effective results in supporting those caring for people with dementia. However, many papers do not separate the results for differing intervention contents or delivery; this needs to be considered when drawing conclusions. There was an overlap among the studies included in the reviews. This suggests a lack of current research on the effectiveness of web-based interventions for people caring for a person with dementia. There was also a lack of consistency in the outcome measures across all papers. Future studies can involve updating research on the effectiveness of these interventions while distinguishing between different intervention types, thus creating guidelines for the use of standardized measures to enable comparisons of intervention effects and improve the scientific quality of the overall research.

Trial Registration: PROSPERO CRD42021241559; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=241559

(J Med Internet Res 2022;24(7):e36727) doi:10.2196/36727

KEYWORDS

dementia; Alzheimer disease; informal; family; carers; caregivers; internet; online; technology; interventions

Introduction

Background

Caring for someone with dementia is challenging emotionally and physically, and carers often need support in this role [1]. This is amplified when the person with dementia is a family member, and the carers are unlikely to have received any formal training in dementia knowledge or how to care for a person with dementia [2]. Supporting informal carers helps not only the carer but also the person with dementia by improving their quality of life and care, thus enabling them to remain at home for longer [1]. Informal carers are people who offer care and support to a person with dementia on a familial or friendship basis. An umbrella review of psychosocial interventions for informal carers of persons with dementia in 2017 identified 13 studies [1]. These included randomized controlled trials (RCTs) of interventions aimed at reducing stress, depression, and other mental health issues as well as physical health problems. They used a combination of educational resources on dementia, practical caring advice, and tips as well as peer support and psychotherapeutic methods to help carers adjust their way of thinking and cope with behavior changes related to dementia. Dickinson et al [1] concluded that multicomponent interventions consisting of educational, social, and therapeutic elements were most effective in improving the well-being of informal carers of persons with dementia. This suggests that carers need advice on caring for persons with dementia as well as for coping and managing their own emotions.

Differentiating between interventions targeting formal (professional) carers and those targeting informal carers is important. Formal carers will often be more experienced in caring for a person with dementia; hence, they will have some prior knowledge and training [2]. The relationship of a person with dementia with a formal carer is different from that with an informal carer, who often knows the person before the diagnosis, suggesting a more emotionally involved relationship. This was demonstrated when carers were asked to describe their reactions to aggressive behaviors from persons with dementia [3]. Informal carers said they devoted more time to the person with dementia and reduced outside contact, whereas formal carers gave practical ways to avoid behavioral episodes that did not isolate them from support systems [3]. This noted difference in approach could benefit from training and understanding of behavior, helping informal carers to consider alternative approaches.

Health care interventions have been moving toward becoming more technology based to combat the rising and unmaintainable health and social care costs because they require fewer staff and are able to reach more people at a similar cost [4]. This move has been accelerated by the COVID-19 pandemic, which forced several services to move to a web-based format, and many people learned how to use technology [5,6]. Technology-based interventions such as those administered on the web or by telephone are time- and cost-effective [1]. They are especially beneficial for carers in rural areas who would ordinarily have to travel to access such services [7]. Rurality may also cause difficulties in accessing the internet [5,6]; however, the number of people in the United Kingdom with access to the internet is consistently increasing [8]. Web-based interventions specifically use the internet, such as an educational website or a peer support forum. Technology-based interventions developed before 2000 were mainly administered by CD-ROM or DVD [4], and for the purpose of this review, they are not included in the definition of web-based intervention.

Web-based interventions that are based solely on education and do not have a *live* element, such as a video call, can be more convenient for busy carers to access at any time. They also require less bandwidth to run [9]; therefore, interventions without video calls will be accessible to more carers. However, not having the engagement and accountability of speaking to another person may reduce adherence to web-based interventions [10] and reduce personal contact.

Technology-based interventions may not be appropriate for all carers because of a range of factors, including age [2]. Spouse or sibling carers of persons with dementia are often older than those caring for a parent with dementia, resulting in a large age range for carers of persons with dementia. However, the number of adults aged ≥75 years in the United Kingdom who use the internet has increased by 26% between 2011 and 2019 [8], with a probable further increase since the COVID-19 pandemic. Given these rapidly evolving developments, it seems timely to review the evidence for web-based support for carers of persons with dementia by conducting an umbrella review of published systematic reviews of web-based interventions.

An umbrella review is a relatively new tool used for evidence synthesis [11]. The method was developed in response to the growing number of systematic reviews being published [11,12]. It is a systematic review of reviews that provides an overview of the information available on a subject [12]. It is often broader in scope than a systematic review and offers a summary that may be useful for policy makers [11,12]. This method of data synthesis was selected after a brief literature search revealed several systematic reviews focusing on the effectiveness of web-based interventions for informal carers of persons with dementia.

Aims of This Review

This umbrella review aimed to synthesize systematic reviews of web-based interventions for informal carers. It will (1) identify types of web-based interventions that have been developed for informal carers of persons with dementia in the community and (2) report on which types are most effective in supporting carers of persons with dementia.

This umbrella review summarizes the topic by using a narrative approach. This may detect gaps in the topic and allow

identification of effective methods for future interventions aimed at supporting informal carers of persons with dementia living in the community and enabling persons with dementia to live at home for longer.

Research Question

The research question for this review was as follows:

1. "What types of web-based interventions have been developed for informal carers of persons with dementia living in the community?"

This was further developed into a subquestion:

2. "What are the types of web-based interventions that are most effective in supporting informal carers of persons with dementia living in the community?"

Methods

Overview

The methods used in this umbrella review were based on the guidelines from the Joanna Briggs Institute (JBI) [12]. The research question, search strategy, and inclusion criteria were developed before conducting the search. This protocol was registered with PROSPERO (International Prospective Register of Systematic Reviews; registration number CRD42021241559).

Inclusion Criteria

The inclusion criteria were generated using the population, concept, and context guide as suggested by JBI [12]:

1. P—The population were informal carers of persons with dementia (eg, family members, friends, and neighbors that identified themselves as carers. There were no restrictions on the amount of time spent caring). Paid carers were excluded from this review.
2. C—The concept involved web-based interventions, which included psychosocial, educational, and therapeutic interventions administered on the web. This review focused specifically on web-based interventions owing to the convenience of being able to access them at any time compared with telephone or face-to-face interventions, which are more time sensitive. Therefore, telephone-only interventions, prerecorded videos, and face-to-face interventions were excluded.
3. C—The context for the review was carers for persons with dementia who were living at home. Carers caring for a person residing in a care home or hospital were excluded from the review, as they would usually include formal carers. Carers in the community, either adjusting to their role or managing ongoing stressors from offering regular care, may require support in adjusting to the role of a carer [13,14]. They may also require web-based interventions that they can access without leaving the house, so that they do not need to make alternative care arrangements to physically access supportive interventions [15]. Carers of people in care homes were excluded as although they continued to care for the person with dementia, their needs were conceptualized as different compared with those managing the care for persons with dementia living in the community.

Only systematic reviews and meta-analyses of web-based interventions were included in this study. We refer to systematic reviews as review papers and papers that they reviewed as study papers. Only reviews in English were included, but date limits were not imposed, as remote interventions before 2000 were mainly administered by CD-ROM or teleconferencing [16] and would not be classified as web-based interventions, and therefore, they were excluded.

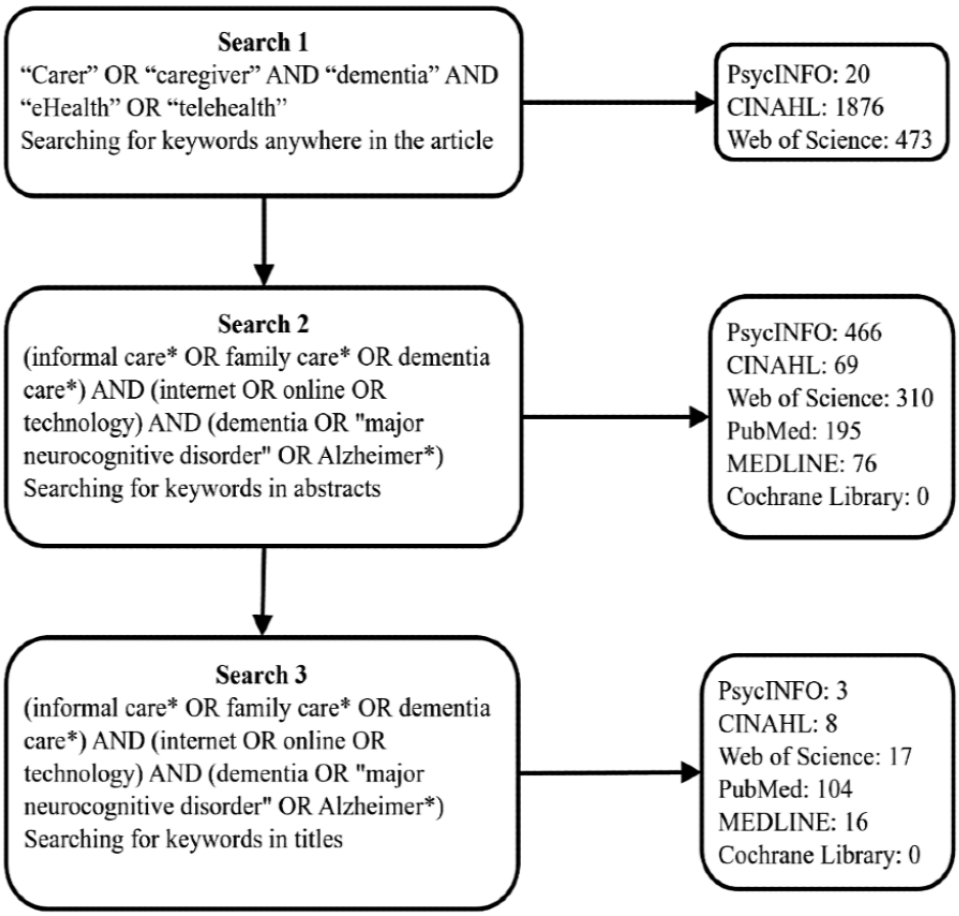
Search Strategy

Overview

The following databases were searched: PsycINFO, Web of Science, CINAHL, MEDLINE, Cochrane Library, and PubMed, as these contain papers from health, social care, and psychological perspectives. The reference lists of review papers were also manually searched, and furthermore, experts were consulted regarding the reviews to be included.

In total, 3 pilot searches were performed to refine the search criteria, focusing on different keywords in different parts of the papers, which provided a better understanding of the literature to create a concise and comprehensive search strategy (Figure 1). This removed irrelevant papers, resulting in a satisfactory number of review papers that addressed the research question.

Figure 1. Keyword search terms and number of results for the 3 database searches.



Search 1

The first search looked for the keywords anywhere in the papers, including the title, abstract, and text (Figure 1). The terms "dementia," "major neurocognitive disorder," and "Alzheimer*" were selected as they would identify papers focusing on a range of dementia subtypes including vascular dementia, dementia with Lewy bodies, and Alzheimer disease. The option to search for systematic reviews only was selected for each database. This search found 2369 review papers with many being irrelevant, containing interventions for carers of conditions other than dementia. This search was conducted only on the first 3 databases because of the large number of papers that were unrelated to the review. The words eHealth and telehealth were found to be more focused on physical health, such as blood pressure monitoring; consequently, they were replaced in the second search with more specific terms, that is, internet, on the web, and technology.

Search 2

The second search included more variations of the search terms used in the first search and focused on locating these only in the abstracts (Figure 1). This resulted in fewer papers (1116 in total); however, many papers focused on conditions other than dementia such as cancer.

Final Search

The third search used the same search terms, but these had to be included in the titles of the review papers, resulting in 148 results from 6 databases (Figure 1). Owing to the number and relevance of the results from the third search, this was the final search strategy used for the review. The strategy was adjusted for each database to use variations of the search that were functional. The final search was conducted on October 1, 2021. A further 3 papers were included based on the suggestion of the second reviewer.

Study Selection

The results were downloaded from the databases and imported into the RefWorks (ProQuest LLC) web-based reference manager. The inbuilt tool was used to remove duplicates, and then, a manual search removed the remaining duplicates that were not identified by RefWorks, which resulted in the removal of 45 papers.

The lead author (BNM) reviewed the titles and abstracts of all 106 review papers to ensure that they met the inclusion criteria, leading to the exclusion of 75 papers. The full texts of 31 papers were assessed by the lead author (BNM), which led to the exclusion of 10 papers. To ensure that the eligibility criteria were applied consistently, coauthors (GW and CL) independently assessed 50% (5/10) of the excluded papers each. These papers were randomly allocated to the coauthors using a random number generator. The inclusion criteria were clear, and there was no disagreement regarding the exclusion of the papers.

Interventions aimed at formal (professional) carers in a care home or hospital setting were excluded, as this review focused

on interventions for informal carers. However, several of the review papers included data collected from both formal and informal carers, which were still included as the exclusion of the papers would result in the loss of relevant data.

Methodological Quality

The quality of the reviews was assessed using the A Measurement Tool to Assess Systematic Reviews (AMSTAR) 2 [17] quality assessment tool. This is an updated version of the AMSTAR tool [18], and it measures the methodological quality of systematic reviews. There are 16 items assessing the inclusion of systematic review methods that are considered high quality (Table 1). The AMSTAR 2 tool is not intended to rate the quality of a review paper as a whole, and it is advised to consider the impact of each individual item to provide a rating of overall confidence in the results [17]. Ratings range from critically low to high and are produced using the web-based assessment tool. The scores are dependent on the answers to 7 critical items presented in Table 1. A high-quality rating indicates <1 critical weakness in a review paper, moderate rating indicates >1 noncritical weakness, low rating indicates 1 critical weakness, and critically low rating indicates >1 critical weakness. All reviews were assessed for quality by the lead author (BNM), and the second and third authors (GW and RS) assessed 29% (6/21) of the papers each, resulting in 57% (12/21) of the reviews being assessed by 2 authors. The fourth author (CL) was consulted to reach a consensus in the case of any disagreement.

Table 1. A Measurement Tool to Assess Systematic Reviews (AMSTAR) 2 questions, responses, and "critical" items.

AMSTAR 2 questions	Responses	Critical domain
Did the research questions and inclusion criteria for the review include the components of PICO[a]?	Yes or no	No
Did the report of the review contain an explicit statement that the review methods were established before the conduct of the review, and did the report justify any significant deviations from the protocol?	Yes or partial yes or no	Yes
Did the review authors explain their selection of the study designs for inclusion in the review?	Yes or no	No
Did the review authors use a comprehensive literature search strategy?	Yes or partial yes or no	Yes
Did the review authors perform study selection in duplicate?	Yes or no	No
Did the review authors perform data extraction in duplicate?	Yes or no	No
Did the review authors provide a list of excluded studies and justify the exclusions?	Yes or partial yes or no	Yes
Did the review authors describe the included studies in adequate detail?	Yes or partial yes or no	No
Did the review authors use a satisfactory technique for assessing RoB[b] in individual studies that were included in the review?	For RCTs[c]—yes or partial yes or no or includes only NRSI[d]; for NRSI—yes or partial yes or no or includes only RCTs	Yes
Did the review authors report on the sources of funding for the studies included in the review?	Yes or no	No
If meta-analysis was performed, did the review authors use appropriate methods for statistical combination of results?	For RCTs—yes or no or no meta-analysis conducted; for NRSI—yes or no or no meta-analysis conducted	Yes
If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?	Yes or no or no meta-analysis conducted	No
Did the review authors account for RoB in primary studies when interpreting or discussing the results of the review?	Yes or no	Yes
Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?	Yes or no	No
If they performed quantitative synthesis, did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?	Yes or no or no meta-analysis conducted	Yes
Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?	Yes or no	No

[a]PICO: population, intervention, control group, outcome.

[b]RoB: risk of bias.

[c]RCT: randomized controlled trial.

[d]NRSI: nonrandomized studies of intervention.

Corrected Covered Area

Review papers on similar topics may include the same studies in their reviews. Overlap of individual studies in systematic reviews can mask a lack of current research on a given topic. If new systematic reviews are being conducted, it can give a false impression of new evidence. To address the extent of overlap, the corrected covered area (CCA) [19] measure was used. The CCA was developed to compare the overlap of studies reviewed in meta-analysis reviews. Although this review was of systematic reviews, the CCA index was used to demonstrate the overlap of studies in the review papers. The CCA was calculated by multiplying the number of index publications by the number of review papers and then subtracting the number of index publications. This was then divided by the frequency of repeated studies. Index publications are the number of primary studies in the review papers, so that they are only counted once, disregarding any repeats in other review papers. A high CCA score indicates a high percentage of overlap.

The CCA measurement has been described as a "promising" measure of overlap; however, it is easily skewed by the inclusion of a single review containing many index publications [20]. To compensate for this potential skew, 5 steps were recommended: create a citation matrix (Multimedia Appendix 1 [15,16,21-39]), calculate the total CCA, calculate the CCA for reviews with high overlap, examine the topic areas for differences in date or samples, and discuss the potential implications of the overlap and report on similarities and differences in outcomes.

Data Extraction

A table for data extraction was created based on the JBI [12] guidelines for conducting umbrella reviews. The data extracted included authors, date, country in which the study was undertaken, number of studies reviewed, population or demographics, sample sizes, intervention details (content and delivery), measures, aims, results, and key findings related to the research question.

Of 21 review papers, 3 (14%) [21-23] included data from both face-to-face interventions and technology-based interventions. The results from these 2 interventions were reported separately in each review, so only the data from the technology-based interventions were extracted (refer to column S in Multimedia Appendix 2 [15,16,21-39]). However, the definition of "technology-based" interventions varied among some of the review papers. For example, a study by Chang [40], which looked at a cognitive behavioral therapy–based intervention that was delivered by video and telephone, was classified as a technology-based intervention in the reviews by Deeken et al [24], Jackson et al [25], Lucero et al [26], and Waller et al [27] but not by Thompson et al [21]. The rationale for Thompson et al [21] classifying the study by Chang [40] as face-to-face was not clear. Consequently, the data from the study by Chang [40] were extracted from the reviews that defined it as a technology-based intervention [24-27].

Of 21 review papers, 3 (14%) included data from both formal and informal carers. Hopwood et al [28] included 40 studies, 3 (8%) of which were a mixture of formal and informal carers. Etxeberria et al [29] primarily incorporated informal carers, but 2 studies included formal carers. Pleasant et al [16] included 9 studies examining formal and informal carers. However, the results of these studies could not be separated from those of informal carers. As only 11.8% (14/119) of studies included both formal and informal carers, they were included in the data synthesis.

To evaluate the interrater reliability of the data extraction, the second and third reviewers (GW and RS) evaluated 57% (12/21) of the total review papers. Any disagreements were resolved by the fourth reviewer (CL).

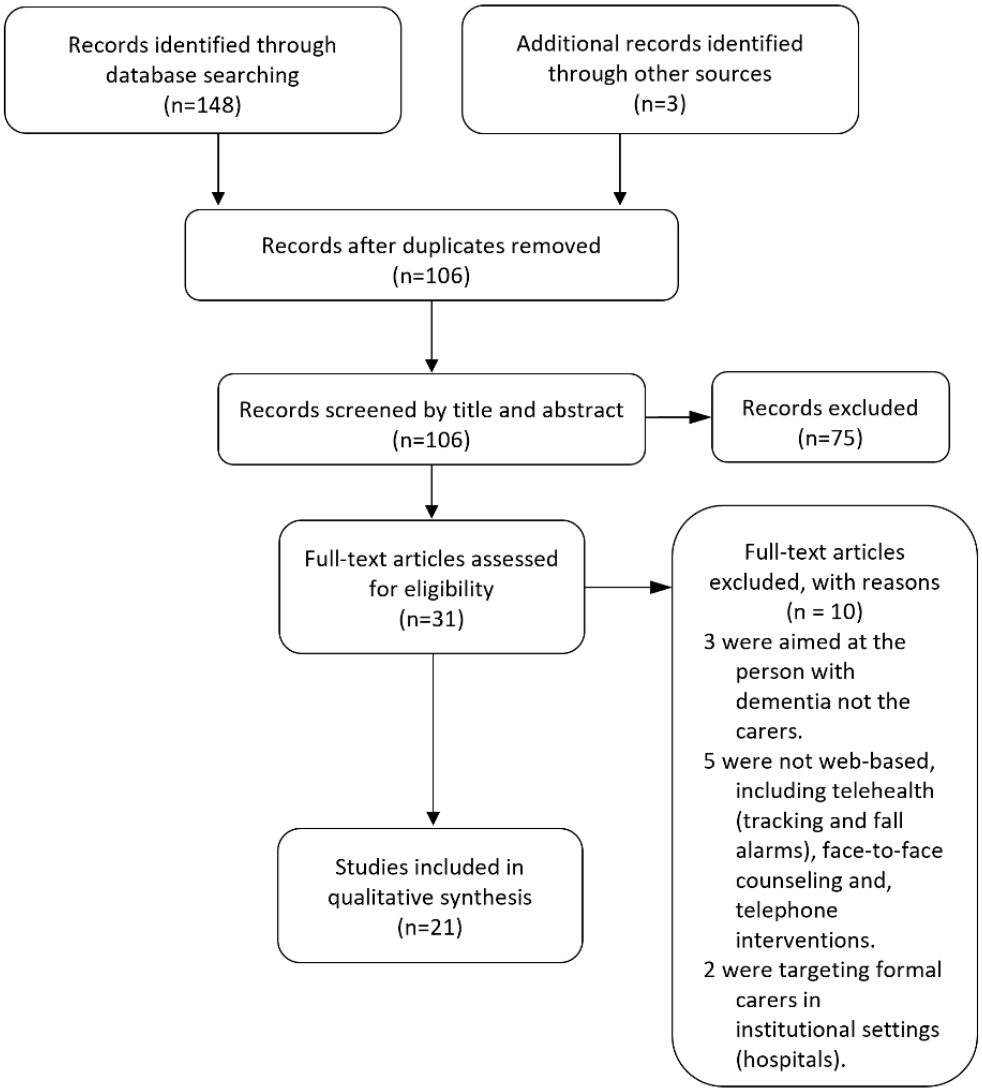
Data Synthesis

In line with the JBI [12] framework for conducting umbrella reviews, the data were presented as a summary of the synthesized results with no further analysis. The extracted data were used to populate a table of study characteristics (Multimedia Appendix 2) and a table summarizing the evidence. The framework states that the summary table should be presented visually, showing effective interventions, mixed results, and no significant improvements. Although the term "major neurocognitive disorder" was introduced in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition [41], to help counter the stigma associated with the term "dementia," research and health care settings continue to use the term dementia. The key terms used in the systematic search included both "dementia" and "major neurocognitive disorder"; however, we found that most of the papers used the term "dementia," and consequently, for the purposes of this review, we use the term "dementia" throughout.

Results

The review process is illustrated in Figure 2. A total of 21 review papers met the inclusion criteria and were included for data synthesis.

Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram of the selection process.



Methodological Quality

After reviewing the papers separately, there was agreement on 83% of the AMSTAR 2 [17] scores. Any disagreements were resolved by the third author. Of the 21 reviews included, 2 (9%) were rated as high quality, 9 (43%) were rated as moderate quality, 2 (9%) were rated as low quality, and 8 (38%) were rated as critically low quality (Multimedia Appendix 2). This suggests that despite the number of papers on a similar topic, only 2 contained the recommended depth. The most common issues were a lack of bias assessment and not having a protocol in place before conducting the search, which were suggested to be because of time constraints.

CCA Scores

The total CCA score showed a 7% overlap, which is considered a moderate overlap; <5% is low overlap, and >15% is high overlap [19]. The CCA matrix (Multimedia Appendix 1) was used to determine reviews with a high overlap so that the CCA scores could be calculated for them. Individual CCA scores were calculated by comparing each of the 21 reviews with each other (Multimedia Appendix 1). The results showed that 26% of the CCA scores were high, 50% were moderate, and 24% were low. The highest overlap was 56%, which was between those reported by Zhao et al [30] and Egan et al [31].

Characteristics of the Included Reviews

The total number of research studies included across the review papers was 119 (50 of these were included more than once). The number of studies included in each review paper ranged from 4 to 40. The research designs varied within each review paper and included mixed methods, RCTs, and pre- and postintervention measures.

Participants

The sample sizes varied from a pre-post intervention study of 4 [28] to an RCT of 1222 participants [32]. The participants included mainly informal carers; 29% (6/21) review papers included a mix of formal and informal carers, and 14% (3/21) review papers also looked at the intervention effects on persons with dementia (Multimedia Appendix 2). There was an umbrella term of “dementia carers” being used to describe formal and informal carers for people with Alzheimer disease, vascular dementia, and early onset dementia; however, 57% (12/21) of the reviews did not describe the type of dementia that the care recipient had. Furthermore, 52% (11/21) reviews included more

detailed information on the participants [15,24,26,27,29,31-35,42], including their nationalities. The countries in which the studies were conducted were reported in 52% (11/21) reviews. In addition, 52% (11/21) reviews reported the gender of carers. Participants were mainly female; however, 1 study [43] that was included in 3 of the reviews specifically looked at male carers [24-26]. Of the 21 reviews, 8 (38%) reviews described age of the carers, which ranged from 18 to 88 years.

Interventions

Interventions were delivered using technology, including web, telephone, DVD, or a combination of these (Multimedia Appendix 2). The duration of the interventions varied from having 24-hour access to websites to 6-minute telephone sessions.

The content also varied; however, psychoeducation, psychotherapy, and social support were the most common types of interventions (Multimedia Appendix 2). Psychoeducational interventions provided information on caring and dementia to the carer. This included information from web-based encyclopedias, practical caring advice, classes with homework, web-based question-and-answer sessions with nurses, and web-based quizzes. The psychotherapy interventions contained elements of cognitive behavioral therapy to help carers manage their emotions and behavior, including cognitive restructuring, relaxation, and "telephone counseling." The interventions offering social support used peer support via web-based groups, individual phone calls, and voice messages, facilitating carers to have contact with people in similar situations. Social support interventions have offered advice and coping strategies; however, this was from peers, and hence, it was more informal. At least one of these 3 intervention types was found to be present in all of the studies, often more than one, which is referred to as a multicomponent intervention.

Outcomes of Review Papers

Outcomes differed between studies, with depression, anxiety, burden, and self-efficacy being most commonly measured. However, the measures used to assess these outcomes varied between the reviews. Waller et al [27] reported 5 different measures for depression (Table 2), and Jackson et al [25] reported 5 measures for burden. This variation in measures made pooling results difficult [31]; however, 7 reviews reported a meta-analysis of the results [15,21,23,24,29,30,36].

Table 2. Depression measures used in the 21 review papers.

Authors	Depression measures	Number of depression measures
Leng et al, 2020 [15]	CES-D[a], BDI-II[b], and PHQ-9[c] questions	3
Deeken et al, 2019 [24]	CES-D, BDI, GDS[d], and BSI[e] depression subscale	4
Hopwood et al, 2018 [28]	CES-D and PHQ-9	2
Boots et al, 2014 [38]	CES-D	1
Egan et al, 2018 [31]	CES-D and BDI-II	2
Jackson et al, 2016 [25]	CES-D, BDI, and GDRS[f]	3
Lucero et al, 2019 [26]	NR[g]	NR
Pleasant et al, 2020 [16]	CES-D, BDI-II, and GDS	3
Scott et al, 2016 [36]	CES-D and BDI-SF[h]	2
Waller et al, 2017 [27]	CES-D, BDI-II, PHQ-9, GDS, and SDS[i]	5
Zhao et al, 2019 [30]	CES-D and BDI-II	2
Etxeberria et al, 2020 [29]	CES-D, BDI, PHQ, and SDS	4
Frias et al, 2020 [22]	CES-D	1
Godwin et al, 2013 [32]	CES-D	1
Kishita et al, 2018 [23]	CES-D and BDI	2
Klimova et al, 2019 [33]	CES-D	1
Parra-Vidales et al, 2017 [37]	CES-D and BDI	2
Powell et al, 2008 [39]	NR	NR
Lee, 2015 [34]	CES-D	1
Thompson et al, 2007 [21]	NR	NR
McKechnie et al, 2014 [35]	CES-D	1

[a]CES-D: Center for Epidemiologic Studies Depression Scale.
[b]BDI-II: Beck Depression Inventory-II.
[c]PHQ-9: Patient Health Questionnaire-9.
[d]GDS: Geriatric Depression Scale.
[e]BSI: Brief Symptom Inventory.
[f]GDRS: Geriatric Depression Rating Scale.
[g]NR: not reported.
[h]BDI-SF: Beck Depression Inventory short form.
[i]SDS: Zung Self-Rating Depression Scale.

Summary of the Evidence

Table 3 is the summary of evidence table recommended by JBI [12]. It shows the 4 most commonly reported outcomes for each review paper. Parra-Vidales et al [37] reported 7 studies but discussed the intervention content of 5. Carer depression improved in 71% (5/7) of the meta-analyses, and anxiety improved significantly in 75% (3/4) of the meta-analyses. The outcomes for caregiver burden were inconsistent; 20% (1/5) reported improvement, 60% (3/5) found no effect or improvements in the control groups, and 20% (1/5) had mixed results. Self-efficacy was investigated in 1 meta-analysis [15], which found positive results. A total of 5 reviews stated that the personalization of interventions was important [15,16,28,38,42]. In total, 3 reviews suggested a need for further research into the intervention effects for specific population groups, such as caring for people with different types of dementia or the nature of the relationships with the person with dementia (spouse, child, and sibling) [24,25,39]. Multicomponent interventions that combined telephone and internet delivery methods with elements of education, psychotherapy, and social support had the largest reported effect sizes [24,25,29,38]. However, most review papers did not differentiate the results for each type of intervention or delivery method.

Table 3. Summary of intervention content, delivery, and results of the 4 most commonly measured outcomes.

Study	Intervention content, n (%)				Intervention delivery, n (%)				Depression[a]	Anxiety[a]	Burden[a]	Self-efficacy[a]
	Educational	Psychotherapeutic	Social	Multicomponent	Telephone	DVD	Computer	Mixed				
Leng et al, 2020 [15][b]	9 (52.9)	1 (5.9)	0 (0)	7 (41.2)	0 (0)	0 (0)	14 (82.4)	3 (17.6)	+	+	–	+
Deeken et al, 2019 [24][b]	6 (18.2)	7 (21.2)	2 (6.1)	18 (54.5)	11 (33.3)	1 (3)	11 (33.3)	10 (30.3)	+	NR[c]	+	NR
Hopwood et al, 2018 [28]	8 (20)	0 (0)	2 (5)	30 (75)	0 (0)	0 (0)	39 (97.5)	1 (2.5)	+	+	+	+
Boots et al, 2014 [38]	3 (25)	0 (0)	0 (0)	9 (75)	0 (0)	0 (0)	9 (75)	3 (25)	+	–	?	+
Egan et al, 2018 [31]	4 (50)	1 (12.5)	0 (0)	3 (37.5)	0 (0)	0 (0)	6 (75)	2 (25)	?	+	–	+
Jackson et al, 2016 [25]	6 (27.3)	3 (13.6)	2 (9.1)	11 (50.0)	13 (59.1)	0 (0)	5 (22.7)	4 (18.2)	?	+	+	+
Lucero et al, 2019 [26]	0 (0)	0 (0)	0 (0)	12 (100)	6 (50)	1 (8.3)	2 (16.7)	3 (25)	?	?	?	?
Pleasant et al, 2020 [16]	18 (94.7)	0 (0)	0 (0)	1 (5.3)	0 (0)	0 (0)	19 (100)	0 (0)	+	+	?	?
Scott et al, 2016 [36][b]	0 (0)	0 (0)	0 (0)	4 (100)	0 (0)	2 (50)	2 (50)	0 (0)	+	NR	NR	NR
Waller et al, 2017 [27]	0 (0)	12 (35.3)	3 (8.8)	19 (55.9)	15 (44.1)	0 (0)	10 (29.4)	9 (26.5)	?	?	?	?
Zhao et al, 2019 [30][b]	2 (33.3)	0 (0)	0 (0)	4 (66.7)	0 (0)	0 (0)	4 (66.7)	2 (33.3)	+	+	–	NR
Etxeberria et al, 2020 [29][b]	4 (44.4)	0 (0)	0 (0)	5 (55.6)	0 (0)	0 (0)	9 (100)	0 (0)	+	–	–	NR
Frias et al, 2020 [22]	4 (50)	4 (50)	0 (0)	0 (0)	3 (37.5)	0 (0)	3 (37.5)	2 (22.2)	?	?	?	+
Godwin et al, 2013 [32]	1 (12.5)	0 (0)	0 (0)	7 (87.5)	4 (50)	0 (0)	3 (37.5)	1 (12.5)	+	+	+	+
Kishita et al, 2018 [23][b]	4 (44.4)	5 (55.6)	0 (0)	0 (0)	4 (44.4)	2 (22.2)	2 (22.2)	1 (11.1)	?	+	?	NR
Klimova et al, 2019 [33]	4 (66.7)	1 (16.7)	0 (0)	1 (16.7)	0 (0)	0 (0)	6 (100)	0 (0)	+	NR	NR	NR
Parra-Vidales et al, 2017 [37]	5 (71.4)	0 (0)	0 (0)	2 (28.6)	0 (0)	0 (0)	6 (85.7)	1 (14.3)	+	+	NR	?
Powell et al, 2008 [39]	0 (0)	2 (13.3)	0 (0)	13 (86.7)	2 (13.3)	0 (0)	11 (73.3)	2 (13.3)	?	?	?	+
Lee, 2015 [34]	0 (0)	0 (0)	5 (100)	0 (0)	0 (0)	1 (20.0)	3 (60.0)	1 (20.0)	+	NR	?	+
Thompson et al, 2007 [21][b]	0 (0)	0 (0)	0 (0)	4 (100)	2 (50)	0 (0)	1 (25)	1 (25)	–	NR	NR	NR
McKechnie et al, 2014 [35]	0 (0)	8 (57.1)	0 (0)	6 (42.9)	0 (0)	1 (7.1)	11 (78.6)	2 (14.3)	?	+	+	?

[a]+ indicates an improvement, ? indicates mixed results, and – indicates no significant improvement or improvements in the control group.

[b]The papers are the results from meta-analyses rather than the results from individual studies.

[c]NR: not reported.

Conclusions of Review Papers

The overall conclusion from the review papers was that technology-based interventions were effective in helping informal carers of persons with dementia [22], but there were difficulties in comparing the effectiveness because of variations in study methods (Multimedia Appendix 3 [15,16,21-39]). Some studies have suggested that rigorous and standardized methods are needed to enable effective comparisons between interventions [30-32]. Low-quality RCTs have been reported, and suggestions for future research included the use of high-quality studies [22,27,29,38].

Another suggestion from the review papers was that the lack of adherence to interventions by carers needs further investigation, with many studies not reporting on adherence, attrition bias, and poor response rates [27,37]. To overcome this, feedback from carers regarding the acceptability of the intervention could be collected. Regarding carer characteristics, 10% (2/21) of the review papers suggested further exploration of the impact of technology-based interventions for carers who look after persons with dementia at different stages of dementia or with differing dementia types [25,28]. It has also been suggested that the long-term effects of interventions should be assessed by collecting data at more time points [16,36] and after booster sessions [16]. Training on how to use the technology may also be useful so that those unfamiliar with it are not discouraged from using the intervention [33].

Findings of This Review

The first objective of this review was to identify the types of interventions that have been developed for informal carers of persons with dementia and to report on their effectiveness. The intervention content was classified into 4 categories: psychoeducational, psychotherapeutic, social, and multicomponent. The delivery of "technology-based interventions" was classified into the categories of telephone, web-based, DVD, and mixed.

The second aim was to report the most effective type of web-based intervention. The findings suggest that multicomponent interventions were the most effective, especially when they used mixed delivery methods, such as telephone and computer. The 3 review papers that included technology-based interventions as well as face-to-face interventions reported that technology-based interventions were just as effective as face-to-face interventions [21-23].

Discussion

Principal Findings

Web-based interventions and services (eHealth) are a rapidly growing area of health care delivery. This study identified and synthesized findings from 21 systematic review papers that examined the effectiveness of web-based interventions to support informal carers of persons with dementia. Web-based interventions were evaluated in 119 studies across the reviews. The delivery methods and intervention content varied between interventions, but the most effective interventions were mixed delivery and mixed content.

Clinical Findings

The findings of this review suggest that technology-based interventions are beneficial in supporting informal carers of persons with dementia. However, several points need to be considered by clinicians before recommending and implementing an intervention.

Several reviews suggested that personalization of interventions to specific carer groups was beneficial, including differentiating between carers caring for patients at different stages of dementia, carers caring for patients with different types of dementia, and carers with different levels of experience [15,16,28,31,38]. Leng et al [15] reported that a questionnaire asking informal carers about their time available to participate and the severity of the dementia of the person they are caring for allowed interventions to be customized for the individual situation.

The delivery method is important to consider. Mixed delivery methods showed greater improvements than web-based, telephone, or DVD-based interventions alone. This could be because of the reported lack of adherence to web-based interventions that do not involve personal contact with the clinicians [27,37]. The personalized contact in mixed delivery interventions could support participants to continue with the intervention using "live" telephone calls by clarifying queries arising from web-based materials and addressing other concerns that arise [7].

Although multiple review papers have analyzed research in this field to date, there has been no umbrella review. One of the unique aspects of an umbrella review is that it considers the overlap of studies included in other review papers. It is thus able to review when data are being reported multiple times and synthesize the data, giving a clearer insight into the effectiveness of, in this paper, the eHealth interventions. Of the 119 studies included in the review, 50 (42%) were reported more than once. The results from the studies by Kajiyama et al [44] and Beauchamp et al [45] were included in 12 of the reviews, which could give an unbalanced and unwarranted emphasis on their findings. An umbrella review considers these factors and offers clinicians an integrated review where these factors are considered. This is the first umbrella review on this topic, and therefore, it has important clinical and research implications.

Strengths and Limitations

The differences in intervention content affected comparisons of data, with many reviews reporting their findings descriptively rather than pooling the diverse statistical data. The quality scores may also have been skewed by the heterogeneous methods in many reviews, resulting in a lack of meta-analyses. A total of 2 of the critical AMSTAR 2 items [17] assessed whether a meta-analysis was conducted; therefore, a negative response to those items affected the results. Further supporting this point, the only 2 reviews to score as high quality contained meta-analyses [15,24]. This may be a limitation of AMSTAR 2, leading to downgrading reviews that were otherwise well conducted but could not conduct a meta-analysis owing to the heterogeneity of the data. AMSTAR 2 was selected over the JBI recommended critical appraisal checklist [12] as it is widely recommended to assess the quality of systematic reviews [46].

The CCA score was used to ascertain the overlap of the studies included in the reviews. A measure of overlap was selected owing to the presence of 1 study [44] in 12 of the reviews [15,16,22,24-26,28,30,31,33,36,37]. It is suggested that CCA can be misrepresentative of the true overlap by the inclusion of several "index publications." The overall moderate score of 7% overlap, yet the high scores of comparisons between individual papers, suggests that the overall score may have been skewed by a high number of "index publications" [20]. The high overlap between several review papers could mean that despite "new" systematic reviews being conducted, there is a false impression of new evidence. This could mask the lack of research on this topic.

A limitation of reviewing the reviews is that the data were limited to what the previous authors reported on. The results were not separated for different intervention types, formal or informal carers, and delivery methods, making it difficult to conclude on the most effective intervention type. Standardized and more scientifically rigorous methods have been suggested to enable better comparison, so development of more effective interventions can occur [30-32].

Demographic data, such as age or dementia type, were not reported in many of the reviews. This could be because the studies themselves did not report the data, thus limiting the review papers. The countries in which the studies were conducted were all high-income countries, showing a potential lack of research in low-income countries. Rural countries may benefit the most from web-based interventions as the support can be provided to carers from anywhere. However, these interventions assume that carers are able to read, have an internet connection, and access a computer.

Conclusions and Recommendations

The findings of this review demonstrated the use of a wide variety of intervention methods and delivery, content, and outcome measures when studying web-based interventions for carers of people with dementia, making comparisons across studies difficult. Web-based interventions showed most positive results for improving carers' depression and anxiety; however, other outcomes were not consistent. Interventions that included psychoeducation, psychosocial, and psychotherapeutic elements were the most effective in improving carer well-being. Tailored interventions for the individual to ensure relevance also improved effectiveness. Nevertheless, the methodological quality of many of the review papers was "critically low," and this needs to be considered when interpreting the results.

Web-based interventions have the potential to reach informal carers of persons with dementia in geographically isolated areas where support may be difficult to access. Further research should aim to distinguish between the types of technology-based intervention content and delivery, thus enabling easier comparison of results. Other suggestions include a focus on specific groups of carers, such as those with different relationships with persons with dementia and those caring for patients with varying stages and types of dementia, which could help personalize the interventions and potentially encourage carers to continue to use the interventions. Research focusing on specific populations or intervention types will facilitate the development of effective web-based interventions in the future.

This is the first umbrella review to examine the effectiveness of technology-based interventions for informal carers of persons with dementia. Previous research has noted the effectiveness of multicomponent interventions and the importance of personalizing interventions. However, the overlap of research on this topic has not been reported previously. The overlap of research on the effectiveness of eHealth interventions for informal carers of persons with dementia may lead to this topic being neglected, and this has consequences for future research.

Acknowledgments

This work is part of a PhD studentship "Adaptation of an e-health intervention 'iSupport' for carers of people living with rare dementias." and is linked to the Rare Dementia Support Impact project (The impact of multicomponent support groups for those living with rare dementias, ES/S010467/1) and is funded jointly by the Economic and Social Research Council (ESRC) and the National Institute for Health Research. ESRC is part of United Kingdom Research and Innovation. The views expressed are those of the authors and not necessarily those of the ESRC, United Kingdom Research and Innovation, the National Institute for Health Research or the Department of Health and Social Care. Rare Dementia Support is generously supported by the National Brain Appeal [47].

Conflicts of Interest

None declared.

Multimedia Appendix 1
Corrected covered area citation matrix and individual overlap percentages between each of the review papers.
[XLSX File (Microsoft Excel File), 21 KB - jmir_v24i7e36727_app1.xlsx]

Multimedia Appendix 2
Characteristics of the review papers.
[DOCX File , 21 KB - jmir_v24i7e36727_app2.docx]

Multimedia Appendix 3
Aims, results, and conclusions of the review papers.
[DOCX File , 20 KB - jmir_v24i7e36727_app3.docx]

References

1. Dickinson C, Dow J, Gibson G, Hayes L, Robalino S, Robinson L. Psychosocial intervention for carers of people with dementia: What components are most effective and when? A systematic review of systematic reviews. Int Psychogeriatr 2017 Jan;29(1):31-43. [doi: 10.1017/S1041610216001447] [Medline: 27666669]
2. Robinson A, Eccleston C, Annear M, Elliott KE, Andrews S, Stirling C, et al. Who knows, who cares? Dementia knowledge among nurses, care workers, and family members of people living with dementia. J Palliat Care 2018 Dec 19;30(3):158-165. [doi: 10.1177/082585971403000305] [Medline: 25265739]
3. Herron RV, Wrathall MA. Putting responsive behaviours in place: examining how formal and informal carers understand the actions of people with dementia. Soc Sci Med 2018 May;204:9-15. [doi: 10.1016/j.socscimed.2018.03.017] [Medline: 29549870]
4. Martínez-Alcalá CI, Pliego-Pastrana P, Rosales-Lagarde A, Lopez-Noguerola JS, Molina-Trinidad EM. Information and communication technologies in the care of the elderly: systematic review of applications aimed at patients with dementia and caregivers. JMIR Rehabil Assist Technol 2016 May 02;3(1):e6 [FREE Full text] [doi: 10.2196/rehab.5226] [Medline: 28582258]
5. Lai J, Widmar NO. Revisiting the digital divide in the COVID-19 era. Appl Econ Perspect Policy 2021 Mar;43(1):458-464 [FREE Full text] [doi: 10.1002/aepp.13104] [Medline: 33230409]
6. De' R, Pandey N, Pal A. Impact of digital surge during Covid-19 pandemic: a viewpoint on research and practice. Int J Inf Manage 2020 Dec;55:102171 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2020.102171] [Medline: 32836633]
7. Ritterband LM, Tate DF. The science of internet interventions. Introduction. Ann Behav Med 2009 Aug;38(1):1-3. [doi: 10.1007/s12160-009-9132-5] [Medline: 19816750]
8. Internet Users, UK 2019. Office of National Statistics. 2019. URL: https://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/bulletins/internetusers/2019 [accessed 2021-11-14]
9. Zhang X, Xu Y, Hu H, Liu Y, Guo Z, Wang Y. Profiling Skype video calls: rate control and video quality. In: Proceedings of the 2012 IEEE INFOCOM. 2012 Presented at: INFOCOM '12; March 25-30, USA; Orlando, FL, USA p. 621-629. [doi: 10.1109/INFCOM.2012.6195805]
10. Wojtowicz M, Day V, McGrath PJ. Predictors of participant retention in a guided online self-help program for university students: prospective cohort study. J Med Internet Res 2013 May 22;15(5):e96 [FREE Full text] [doi: 10.2196/jmir.2323] [Medline: 23697614]
11. Pollock M, Fernandes RM, Becker LA, Pieper D, Hartling L. Chapter V: Overviews of reviews. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions version 6.3. London, UK: Cochrane Collaboration; 2022.
12. Aromataris E, Fernandez R, Godfrey C, Holly C, Khalil H, Tungpunkom P. Chapter 10: Umbrella reviews. In: Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis. Adelaide, Australia: Joanna Briggs Institute; 2020.
13. de Vugt ME, Verhey FR. The impact of early dementia diagnosis and intervention on informal caregivers. Prog Neurobiol 2013 Nov;110:54-62. [doi: 10.1016/j.pneurobio.2013.04.005] [Medline: 23689068]
14. Afram B, Verbeek H, Bleijlevens MH, Hamers JP. Needs of informal caregivers during transition from home towards institutional care in dementia: a systematic review of qualitative studies. Int Psychogeriatr 2015 Jun;27(6):891-902. [doi: 10.1017/S1041610214002154] [Medline: 25287064]
15. Leng M, Zhao Y, Xiao H, Li C, Wang Z. Internet-based supportive interventions for family caregivers of people with dementia: systematic review and meta-analysis. J Med Internet Res 2020 Sep 09;22(9):e19468 [FREE Full text] [doi: 10.2196/19468] [Medline: 32902388]
16. Pleasant M, Molinari V, Dobbs D, Meng H, Hyer K. Effectiveness of online dementia caregivers training programs: a systematic review. Geriatr Nurs 2020;41(6):921-935. [doi: 10.1016/j.gerinurse.2020.07.004] [Medline: 32703628]
17. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017 Sep 21;358:j4008 [FREE Full text] [doi: 10.1136/bmj.j4008] [Medline: 28935701]
18. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol 2007 Feb 15;7:10 [FREE Full text] [doi: 10.1186/1471-2288-7-10] [Medline: 17302989]
19. Pieper D, Antoine SL, Mathes T, Neugebauer EA, Eikermann M. Systematic review finds overlapping reviews were not mentioned in every other overview. J Clin Epidemiol 2014 Apr;67(4):368-375. [doi: 10.1016/j.jclinepi.2013.11.007] [Medline: 24581293]
20. Hennessy EA, Johnson BT. Examining overlap of included studies in meta-reviews: guidance for using the corrected covered area index. Res Synth Methods 2020 Jan;11(1):134-145 [FREE Full text] [doi: 10.1002/jrsm.1390] [Medline: 31823513]

21. Thompson CA, Spilsbury K, Hall J, Birks Y, Barnes C, Adamson J. Systematic review of information and support interventions for caregivers of people with dementia. BMC Geriatr 2007 Jul 27;7:18 [FREE Full text] [doi: 10.1186/1471-2318-7-18] [Medline: 17662119]
22. Frias CE, Garcia-Pascual M, Montoro M, Ribas N, Risco E, Zabalegui A. Effectiveness of a psychoeducational intervention for caregivers of people with dementia with regard to burden, anxiety and depression: a systematic review. J Adv Nurs 2020 Mar;76(3):787-802. [doi: 10.1111/jan.14286] [Medline: 31808211]
23. Kishita N, Hammond L, Dietrich CM, Mioshi E. Which interventions work for dementia family carers?: an updated systematic review of randomized controlled trials of carer interventions. Int Psychogeriatr 2018 Nov;30(11):1679-1696. [doi: 10.1017/S1041610218000947] [Medline: 30017008]
24. Deeken F, Rezo A, Hinz M, Discher R, Rapp MA. Evaluation of technology-based interventions for informal caregivers of patients with dementia-a meta-analysis of randomized controlled trials. Am J Geriatr Psychiatry 2019 Apr;27(4):426-445. [doi: 10.1016/j.jagp.2018.12.003] [Medline: 30642650]
25. Jackson D, Roberts G, Wu ML, Ford R, Doyle C. A systematic review of the effect of telephone, internet or combined support for carers of people living with Alzheimer's, vascular or mixed dementia in the community. Arch Gerontol Geriatr 2016;66:218-236. [doi: 10.1016/j.archger.2016.06.013] [Medline: 27372903]
26. Lucero RJ, Fehlberg EA, Patel AG, Bjarnardottir RI, Williams R, Lee K, et al. The effects of information and communication technologies on informal caregivers of persons living with dementia: a systematic review. Alzheimers Dement (N Y) 2019 Dec 31;5:1-12 [FREE Full text] [doi: 10.1016/j.trci.2018.11.003] [Medline: 30623020]
27. Waller A, Dilworth S, Mansfield E, Sanson-Fisher R. Computer and telephone delivered interventions to support caregivers of people with dementia: a systematic review of research output and quality. BMC Geriatr 2017 Nov 16;17(1):265 [FREE Full text] [doi: 10.1186/s12877-017-0654-6] [Medline: 29145806]
28. Hopwood J, Walker N, McDonagh L, Rait G, Walters K, Iliffe S, et al. Internet-based interventions aimed at supporting family caregivers of people with dementia: systematic review. J Med Internet Res 2018 Jun 12;20(6):e216 [FREE Full text] [doi: 10.2196/jmir.9548] [Medline: 29895512]
29. Etxeberria I, Salaberria K, Gorostiaga A. Online support for family caregivers of people with dementia: a systematic review and meta-analysis of RCTs and quasi-experimental studies. Aging Ment Health 2021 Jul;25(7):1165-1180. [doi: 10.1080/13607863.2020.1758900] [Medline: 32363901]
30. Zhao Y, Feng H, Hu M, Hu H, Li H, Ning H, et al. Web-based interventions to improve mental health in home caregivers of people with dementia: meta-analysis. J Med Internet Res 2019 May 06;21(5):e13415 [FREE Full text] [doi: 10.2196/13415] [Medline: 31066680]
31. Egan KJ, Pinto-Bruno Á, Bighelli I, Berg-Weger M, van Straten A, Albanese E, et al. Online training and support programs designed to improve mental health and reduce burden among caregivers of people with dementia: a systematic review. J Am Med Dir Assoc 2018 Mar;19(3):200-6.e1. [doi: 10.1016/j.jamda.2017.10.023] [Medline: 29306605]
32. Godwin KM, Mills WL, Anderson JA, Kunik ME. Technology-driven interventions for caregivers of persons with dementia: a systematic review. Am J Alzheimers Dis Other Demen 2013 May;28(3):216-222 [FREE Full text] [doi: 10.1177/1533317513481091] [Medline: 23528881]
33. Klimova B, Valis M, Kuca K, Masopust J. E-learning as valuable caregivers' support for people with dementia - a systematic review. BMC Health Serv Res 2019 Nov 01;19(1):781 [FREE Full text] [doi: 10.1186/s12913-019-4641-9] [Medline: 31676005]
34. Lee E. Do technology-based support groups reduce care burden among dementia caregivers? A review. J Evid Inf Soc Work 2015;12(5):474-487. [doi: 10.1080/15433714.2014.930362] [Medline: 25794367]
35. McKechnie V, Barker C, Stott J. Effectiveness of computer-mediated interventions for informal carers of people with dementia-a systematic review. Int Psychogeriatr 2014 Oct;26(10):1619-1637 [FREE Full text] [doi: 10.1017/S1041610214001045] [Medline: 24989249]
36. Scott JL, Dawkins S, Quinn MG, Sanderson K, Elliott KE, Stirling C, et al. Caring for the carer: a systematic review of pure technology-based cognitive behavioral therapy (TB-CBT) interventions for dementia carers. Aging Ment Health 2016 Aug;20(8):793-803. [doi: 10.1080/13607863.2015.1040724] [Medline: 25978672]
37. Parra-Vidales E, Soto-Pérez F, Perea-Bartolomé MV, Franco-Martín MA, Muñoz-Sánchez JL. Online interventions for caregivers of people with dementia: a systematic review. Actas Esp Psiquiatr 2017 May;45(3):116-126 [FREE Full text] [Medline: 28594057]
38. Boots LM, de Vugt ME, van Knippenberg RJ, Kempen GI, Verhey FR. A systematic review of Internet-based supportive interventions for caregivers of patients with dementia. Int J Geriatr Psychiatry 2014 Apr;29(4):331-344. [doi: 10.1002/gps.4016] [Medline: 23963684]
39. Powell J, Chiu T, Eysenbach G. A systematic review of networked technologies supporting carers of people with dementia. J Telemed Telecare 2008;14(3):154-156. [doi: 10.1258/jtt.2008.003018] [Medline: 18430288]
40. Chang BL. Cognitive-behavioral intervention for homebound caregivers of persons with dementia. Nurs Res 1999;48(3):173-182. [doi: 10.1097/00006199-199905000-00007] [Medline: 10337848]
41. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders Text Revision (DSM-5-TR). 5th edition. Washington, DC, USA: American Psychiatric Association; 2013.

42. Christie HL, Bartels SL, Boots LM, Tange HJ, Verhey FJ, de Vugt ME. A systematic review on the implementation of eHealth interventions for informal caregivers of people with dementia. Internet Interv 2018 Sep;13:51-59 [FREE Full text] [doi: 10.1016/j.invent.2018.07.002] [Medline: 30206519]
43. Gant JR, Steffen AM, Lauderdale SA. Comparative outcomes of two distance-based interventions for male caregivers of family members with dementia. Am J Alzheimers Dis Other Demen 2007;22(2):120-128 [FREE Full text] [doi: 10.1177/1533317506298880] [Medline: 17545139]
44. Kajiyama B, Thompson LW, Eto-Iwase T, Yamashita M, Di Mario J, Marian Tzuang Y, et al. Exploring the effectiveness of an Internet-based program for reducing caregiver distress using the iCare Stress Management e-Training Program. Aging Ment Health 2013;17(5):544-554 [FREE Full text] [doi: 10.1080/13607863.2013.775641] [Medline: 23461355]
45. Beauchamp N, Irvine AB, Seeley J, Johnson B. Worksite-based Internet multimedia program for family caregivers of persons with dementia. Gerontologist 2005 Dec;45(6):793-801. [doi: 10.1093/geront/45.6.793] [Medline: 16326661]
46. Ma LL, Wang YY, Yang ZH, Huang D, Weng H, Zeng XT. Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better? Mil Med Res 2020 Feb 29;7(1):7 [FREE Full text] [doi: 10.1186/s40779-020-00238-8] [Medline: 32111253]
47. The National Brain Appeal. URL: https://www.nationalbrainappeal.org/ [accessed 2022-07-12]

Abbreviations

AMSTAR: A Measurement Tool to Assess Systematic Reviews
CCA: corrected covered area
ESRC: Economic and Social Research Council
JBI: Joanna Briggs Institute
RCT: randomized controlled trial

Edited by M Focsa; submitted 24.01.22; peer-reviewed by AS Rigaud, N Davies; comments to author 09.03.22; revised version received 25.05.22; accepted 28.05.22; published 22.07.22.

Please cite as:
Naunton Morgan B, Windle G, Sharp R, Lamers C
eHealth and Web-Based Interventions for Informal Carers of People With Dementia in the Community: Umbrella Review
J Med Internet Res 2022;24(7):e36727
URL: https://www.jmir.org/2022/7/e36727
doi:10.2196/36727
PMID:35867388

Review

Implementation of Web-Based Psychosocial Interventions for Adults With Acquired Brain Injury and Their Caregivers: Systematic Review

Melissa Miao[1], BAppSc(Hons); Rachael Rietdijk[2], BAppSc(Hons), PhD; Melissa Brunner[2], BAppSc, MHlthSc, PhD; Deborah Debono[1], RN, BA(Hons), GCTLHEd, PhD; Leanne Togher[2], BAppSc, PhD; Emma Power[1], BAppSc(Hons), PhD

[1]University of Technology Sydney, Sydney, Australia

[2]University of Sydney, Sydney, Australia

Corresponding Author:
Melissa Miao, BAppSc(Hons)
University of Technology Sydney
PO Box 123 Broadway, Ultimo
Sydney, 2007
Australia
Phone: 61 2 9514 1448
Email: melissa.miao@uts.edu.au

Abstract

Background: More than 135 million people worldwide live with acquired brain injury (ABI) and its many psychosocial sequelae. This growing global burden necessitates scalable rehabilitation services. Despite demonstrated potential to increase the accessibility and scalability of psychosocial supports, digital health interventions are challenging to implement and sustain. The Nonadoption, Abandonment, Scale-Up, Spread, and Sustainability (NASSS) framework can offer developers and researchers a comprehensive overview of considerations to implement, scale, and sustain digital health interventions.

Objective: This systematic review identified published, peer-reviewed primary evidence of implementation outcomes, strategies, and factors for web-based psychosocial interventions targeting either adults with ABI or their formal or informal caregivers; evaluated and summarized this evidence; synthesized qualitative and quantitative implementation data according to the NASSS framework; and provided recommendations for future implementation. Results were compared with 3 hypotheses which state that *complexity* (dynamic, unpredictable, and poorly characterized factors) in most or all NASSS domains increases likelihood of implementation failure; success is achievable, but difficult with many *complicated* domains (containing multiple interacting factors); and *simplicity* (straightforward, predictable, and few factors) in most or all domains increases the likelihood of success.

Methods: From a comprehensive search of MEDLINE, EMBASE, PsycINFO, CINAHL, Scopus, speechBITE, and neuroBITE, we reviewed primary implementation evidence from January 2008 to June 2020. For web-based psychosocial interventions delivered via standard desktop computer, mobile phone, tablet, television, and virtual reality devices to adults with ABI or their formal or informal caregivers, we extracted intervention characteristics, stakeholder involvement, implementation scope and outcomes, study design and quality, and implementation data. Implementation data were both narratively synthesized and descriptively quantified across all 7 domains (condition, technology, value proposition, adopters, organization, wider system, and their interaction over time) and all subdomains of the NASSS framework. Study quality and risk of bias were assessed using the 2018 Mixed Methods Appraisal Tool.

Results: We identified 60 peer-reviewed studies from 12 countries, including 5723 adults with ABI, 1920 carers, and 50 health care staff. The findings aligned with all 3 hypotheses.

Conclusions: Although studies were of low methodological quality and insufficient number to statistically test relationships, the results appeared consistent with recommendations to reduce complexity as much as possible to facilitate implementation. Although studies excluded individuals with a range of comorbidities and sociocultural challenges, such simplification of NASSS domain 1 may have been necessary to advance intervention value propositions (domain 3). However, to create equitable digital health solutions that can be successfully implemented in real-world settings, it is recommended that developers involve people with ABI, their close others, and health care staff in addressing complexities in domains 2 to 7 from the earliest intervention design stages.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42020186387; https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020186387

International Registered Report Identifier (IRRID): RR2-10.1177/20552076211035988

(J Med Internet Res 2022;24(7):e38100) doi:10.2196/38100

KEYWORDS

complexity; implementation science; internet interventions; acquired brain injury; delivery of health care; caregivers; digital health; psychosocial interventions; psychosocial; brain injury; mobile phone

Introduction

Background

More than 135 million people worldwide live with acquired brain injuries (ABIs), such as stroke and traumatic brain injury (TBI) [1]. The number of people with ABI is projected to grow [2], increasing global need for rehabilitation services [1,2], including supports to manage the complex and ongoing psychosocial impact of ABI on relationships [3,4], mental health [5,6], and employment [7,8]. For these rehabilitation services to be provided at scale, they must be effectively integrated into health care systems [1].

Longstanding challenges in the implementation of evidence-based care have led to the emergence of implementation science research [9]. This includes a specific focus on digital health implementation [10,11]. Despite demonstrated potential to increase the accessibility and scalability of psychosocial supports [12,13], digital health interventions are challenging to implement and sustain [11,14,15]. Current evidence indicates that digital health implementation challenges are predominantly organizational, systemic, and sociotechnical in nature, including interrelated challenges of resources, workflows, interoperability, and legislation [10,15,16]. Therefore, understanding and addressing these challenges require a comprehensive, complexity-based approach, in which the complex, adaptive nature of health care systems, actors, and technologies, as well as the interactions between them, are recognized [17-19].

From this complexity paradigm [18-20], the Nonadoption, Abandonment, Scale-up, Spread, and Sustainability (NASSS) framework [17] offers developers, practitioners, and researchers a comprehensive synthesis of considerations to implement, scale, and sustain digital health interventions, to ensure that critically important systemic and organizational factors are not overlooked. The NASSS framework includes 7 domains of digital health implementation: condition, technology, value proposition, adopters, organization, wider system, and their interaction over time [17]. Each domain includes multiple subdomains, with published definitions of how each specific subdomain can be made *simple*, *complicated*, or *complex* [17]. A complexity paradigm has not yet been adopted for digital interventions targeting ABI, despite both the prevalence of this condition [1] and the value of a condition-specific focus from both theoretical [20] and stakeholder perspectives [21].

To date, the NASSS framework has been used to narratively synthesize digital health implementation findings from informal care [22], mixed home care [23], and video consultations [24] in various populations. However, it has not yet been used to underpin deductive extraction and analysis of qualitative data [25] or quantitative analyses in relation to current hypotheses concerning the potential role of complexity in implementation success [20]. Digital health implementation reviews to date have also relied on generic implementation frameworks [26-28], despite their poor fit to digital health [29]. There is therefore a need to examine existing implementation evidence specific to digital health, ABI, and its psychosocial sequelae, and to do so using a comprehensive framework that acknowledges the complexity of implementing, scaling, and sustaining digital health interventions in real-world settings, if we are to enable these interventions to succeed at a scale that can reach and support current and future global needs.

Aims

Based on a previously published protocol [30], the aims of this review were as follows:

1. Identify, evaluate, and summarize the strength and nature of implementation evidence for web-based psychosocial interventions targeting either people with ABI or their caregivers or both.

2. Synthesize qualitative and quantitative implementation data according to the NASSS framework.

3. Provide recommendations for future implementation based on this synthesis.

A subsequently introduced aim was as follows:

4. Compare findings with 3 hypotheses concerning the NASSS framework [20], which state:

- *Hypothesis 1: "If most or all of the domains can be classified as simple, an intervention is likely to be easy to implement and to be achieved on time and within budget"*;
- *Hypothesis 2: "If many domains are classified as complicated, the intervention will be achievable but difficult, and likely to exceed its timescale and budget";*
- *Hypothesis 3: "If multiple domains are complex, the chances of the intervention succeeding at all are limited."*

Methods

Review Registration and Protocol

This systematic review was prospectively registered in PROSPERO (International Prospective Register of Systematic Reviews; CRD42020186387) [31]. A published protocol [30], including the search strategy and selection criteria, was developed a priori. Subsequent protocol adjustments, with

rationales, are reported in accordance with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines [32].

Search Strategy and Selection Criteria

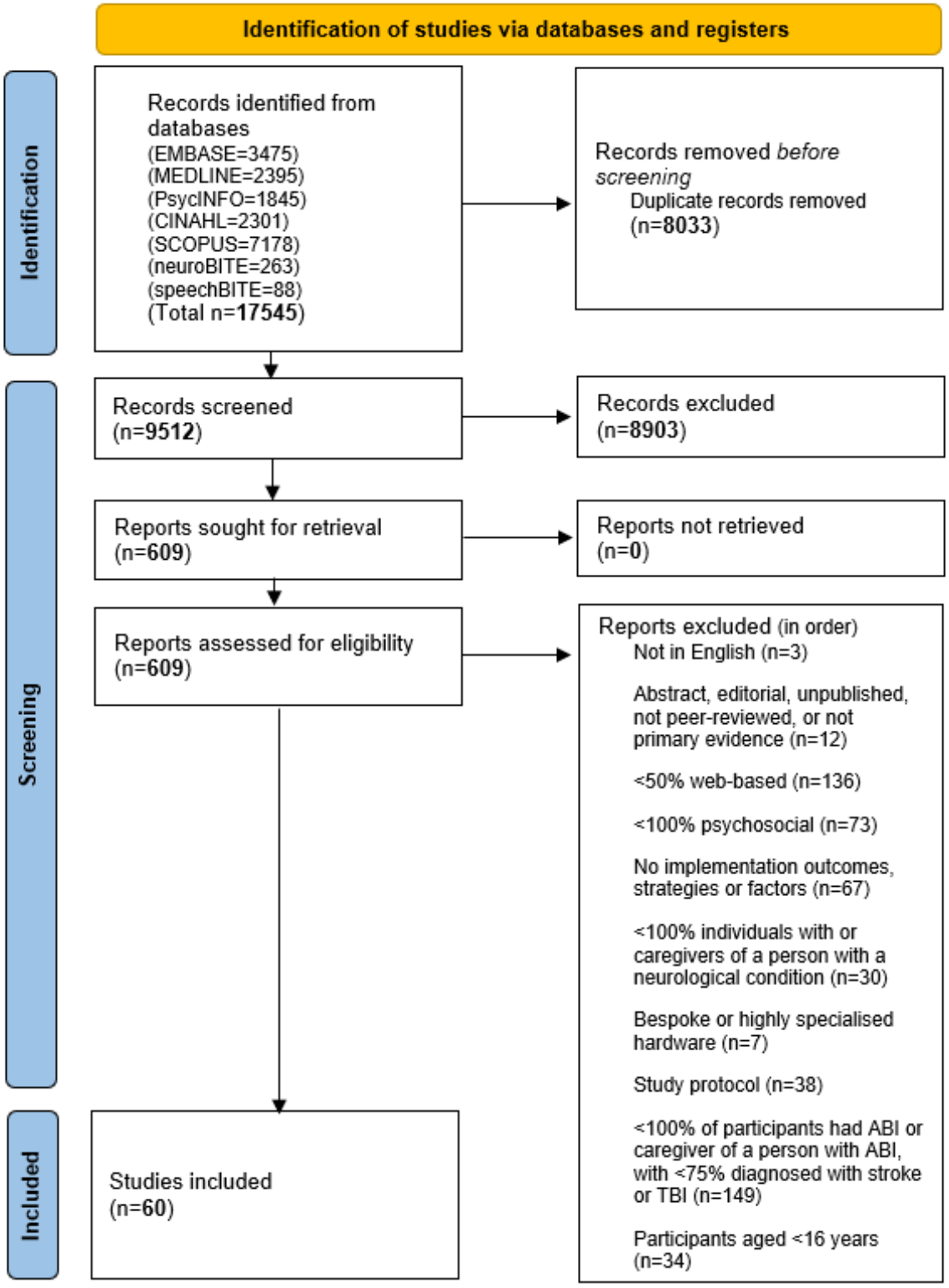
A comprehensive search of 7 databases (MEDLINE, EMBASE, PsycINFO, CINAHL, Scopus, speechBITE, and neuroBITE) was conducted in mid-June 2020 as per the published protocol [30]. The original Population, Intervention, Comparison, Outcome, Study design–based search encompassed multiple neurological conditions (search strategy is available in Multimedia Appendix 1) and returned 17,545 results (refer to Figure 1 for PRISMA flow diagram). After removing duplicates, a total of 9512 titles and abstracts were independently screened using Covidence (Veritas Health Innovation) software [33] by 2 authors (MM and either MB, RR, EP, or DD), applying exclusion criteria in hierarchical order (as listed in Figure 1). There was 96.4% (9170/9512) agreement at the title and abstract level (ie, 3.6% disagreement, or 342/9512 conflicts, including agreed exclusions for conflicting reasons). Conflicts were resolved through consensus discussion by at least 3 authors. A total of 609 records were screened at the full-text level. Due to the high yield of full texts, a pragmatic protocol adjustment was required. Therefore, all full texts were screened by the first author (MM), and a second author (RR, EP, DD, or MB) independently screened 25.1% (153/609) rather than 100% of full texts. There was 82.4% (126/153) agreement in this quarter of full texts (ie, 17.6% disagreement, or 27/153 conflicts, including agreed exclusions for conflicting reasons). The conflicts were discussed by at least 3 authors and resolved by consensus. The team agreed that the reliability of the screening process was adequate for the first author to proceed independently.

Given the high yield of full texts, additional records were not sought as originally planned in the protocol [30]. Instead, additional criteria were selected to increase the review's clinical relevance to our own implementation of web-based psychosocial interventions delivered via standard desktop computers and smart devices to adults with ABI and their communication partners [34]. These narrowed exclusion criteria were introduced in the following hierarchical order (Figure 1):

1. Less than 50% of the web-based intervention was delivered remotely; that is, web-based interventions accessed in a laboratory or clinic were excluded. For example, although Connor et al [34] examined a web-based brain training game, the study was excluded because it focused on in-person delivery on-site, accompanied by face-to-face treatment by a speech-language pathologist.
2. Less than 100% of the intervention was psychosocial in nature, that is, providing cognitive, behavioral, educational, communicational, or supportive care to both the person with the condition or their caregivers. Therefore, interventions with physical rehabilitation (eg, exercise programs, or physical therapy) or health informatics (eg, symptom monitoring, interprofessional communication, or care planning) components were excluded.
3. Less than 100% of participants were diagnosed with a neurological condition or the caregiver of such a person or the results of participants meeting this criterion could not be extracted.
4. The intervention required bespoke or highly specialized hardware beyond standard desktop computer, television, mobile phone, tablet, or virtual reality devices.
5. The record was a study protocol.
6. Less than 100% of intervention recipients were people with ABI or their caregivers, with <75% of the population having had a stroke or TBI or the caregiver of someone with these conditions.
7. Participants were aged <16 years.

The refinement in focus from neurological conditions in criterion 3 to the condition of ABI in criterion 6 aimed to reduce complexity in the first domain of the NASSS framework [20] by "scaling back on the kinds of illness or condition for which the technology is claimed to be useful." It was also introduced to reflect stakeholders' prioritization of the condition of ABI compared with other NASSS domains [21].

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram. ABI: acquired brain injury; TBI: traumatic brain injury.



Extraction

In total, 60 records were included for extraction against the NASSS framework [17]. Due to the high yield requiring in-depth application of the NASSS framework, a pragmatic deviation was required. Therefore, a second author (RR, MB, or DD) checked 25% (15/60) of extractions by the first author (MM) rather than independently extracting all full texts. Additional details were added or changes made in 1.18% (26/2205) of fields, confirmed via written consensus between the 2 rating authors and a third rater if necessary. To ensure consistency, the first author (MM) used a standardized extraction form (Multimedia Appendix 2 [17,36-38]), which included embedded logic via REDCap (Research Electronic Data Capture; the REDCap Consortium) [39], and the published definitions of (1) *complex*, *complicated*, or *simple* for each subdomain of the

NASSS framework [17]; (2) each implementation outcome [36]; and (3) each question in the critical appraisal tool [37]. Text was extracted verbatim, with minimal paraphrasing as required for context.

The data extraction form drafted in our study protocol [30] was updated (Multimedia Appendix 2) to reflect the refined exclusion criteria. New extraction items were also added from a published taxonomy of digital health intervention features [38] to both consistently capture the diversity of interventions and technologies and incorporate implementation considerations identified by stakeholders in a concurrent study [21]. These included the order in which intervention contents were presented and the potential benefit of peer interaction. As described in our protocol [30], study quality, including sampling and nonresponse bias as applicable, was assessed across various study designs using the Mixed Methods Appraisal Tool (MMAT) [37]. Extraction was completed successively via REDCap, ensuring blinding to any emerging patterns until data from all 60 records had been extracted.

Analysis

The high yield of this review enabled the quantification of complexity using descriptive statistics. The numbers of *complex*, *complicated*, and *simple* subdomains and domains and those containing *no information* were each subtotaled. Subdomains were classified according to their published definitions [17] as part of the extraction process (Multimedia Appendix 2). As no domain can be simpler than its constituent parts, each domain for each record was operationally classified according to the most complex subdomain present within that domain (Table 1). Implementation *success* or *failure* for each study was defined as whether the authors succeeded in achieving their specific implementation aims. *Can't tell* was selected (Multimedia Appendix 2) for ambiguous implementation outcomes, such as inconclusive or insufficiently reported implementation results or conflicting implementation and effectiveness results. Such records were excluded, resulting in 75% (45/60) of the records being included in the descriptive analysis of complexity.

In accordance with our protocol [30], all quantitative results were analyzed in REDCap and Microsoft Excel (Microsoft Corporation) using descriptive statistics, and all qualitative results were narratively synthesized according to the NASSS framework [17].

Table 1. Operational classification of domain complexity according to subdomain complexity.

Domain complexity	General definition [17]	Subdomain complexity
Complex	"Dynamic, unpredictable, not easily disaggregated into constituent components"	Complex only; complex and complicated; or complex, complicated, and simple
Complicated	"Multiple interacting components or issues"	Complicated only or complicated and simple
Simple	"Straightforward, predictable, few components"	Simple only

Results

Note Regarding Style

Due to the high number of citations per descriptor, only essential in-text citations are provided. An appended Microsoft Excel spreadsheet containing all bibliographic and extracted data is provided as Multimedia Appendix 3.

Implementation Evidence

More than two-thirds (41/60, 68%) of the reviewed studies were published in 2016 or later (Figure 2). Studies originated from 12 countries (Figure 3), with 3% (2/60) involving international collaboration [40,41]. The most studies were conducted in the United States (21/60, 35%), Australia (16/60, 27%), and Canada (6/60, 10%).

Figure 2. Publication year of included studies.

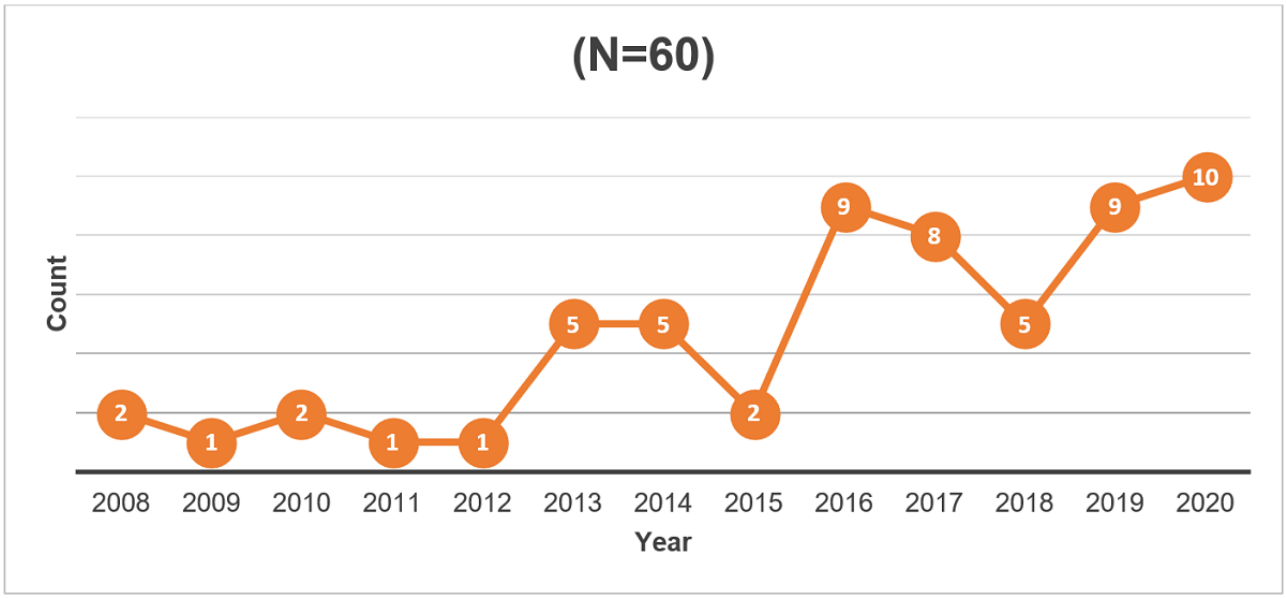
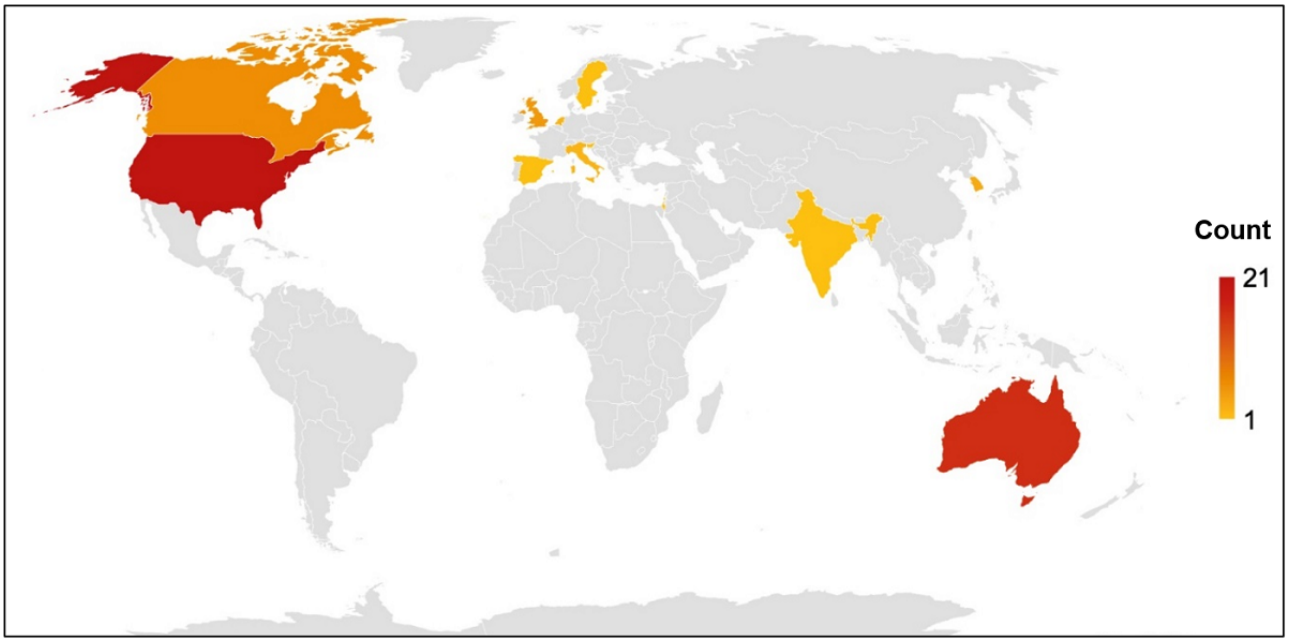


Figure 3. Country of origin of included studies.



Overall, 13% (8/60) of studies used an implementation framework or theory. A generic framework was used in less than half (3/8, 38%) of these studies; Pitt et al [42,43] reported that their intervention had been developed according to a guide for complex intervention development that was not specific to digital health, while Rietdijk et al [44] referred to a generic framework for feasibility to inform study design. Almost twice as many (5/8, 63%) referred to a framework specific to the development of digital interventions, including web-based education [45], web-based programs [46], and user-centered design [47-49].

The most frequently used study design was mixed methods (22/60, 37%), followed by quantitative descriptive (17/60, 28%), quantitative nonrandomized trials (11/60, 18%), quantitative randomized controlled trials (RCTs; 8/60, 13%), and qualitative research (2/60, 3%). The MMAT definition for mixed methods inherently requires mixed methods studies to meet all 5 quality criteria. However, given the reviewed studies that used both qualitative and quantitative methods were typically of poor quality (refer to *Study Quality* section), a deviation was made in which the quality criteria for *mixed methods* were used to describe how these studies fell short of the definition, rather than immediately classifying all these studies as *Other* (Multimedia Appendix 2).

Among hybrid effectiveness implementation studies [50], Type 2 hybrids were the most common (22/60, 37%), followed by Type 1 (17/60, 28%), and Type 3 (4/60, 7%) [51-54]. Of the 60 studies reviewed, 3 (5%) were qualitative studies of implementation [42,55,56], and 14 (23%) used other nonhybrid designs. According to definitions by Proctor et al [36], the most frequently collected implementation measure was feasibility (37/60, 62%), followed by adherence or fidelity (27/60, 45%),

satisfaction (22/60, 37%), usability (19/60, 32%), and acceptability (17/60, 28%). Appropriateness (9/60, 15%), cost-effectiveness (2/60, 3%), and other measures (5/60, 8%) were less frequently observed.

In terms of implementation outcomes, most studies (37/60, 62%) achieved the investigators' specific implementation aims, while a minority (8/60, 13%) failed to do so (Multimedia Appendix 3). There were ambiguous or unsubstantiated implementation outcomes in a quarter (15/60, 25%) of the studies reviewed.

Study Quality

Overall, study quality, as assessed using the MMAT, was low. A quarter of the reviewed studies (15/60, 25%) studies passed all 5 questions in the MMAT, 22% (13/60) of studies passed 4 questions, and 20% (12/60) of studies passed 3 questions. Overall, 12% (7/60) of studies passed 2 questions and 3% (2/60) of studies passed 1 question. Due to a lack of justification for the use of mixed methods and an absence of rigor in qualitative methods, 18% (11/60) of studies did not pass any of the 5 questions in the MMAT. There were clear research questions in 62% (37/60) of included studies, and 65% (39/60) appeared to have clear alignment between implementation aims and outcome measures. Finally, it was noted that studies often excluded individuals based on a range of comorbidities and sociocultural factors typically associated with ABI, and focused on the chronic stage of recovery (refer to *Domain 1: Condition* section). This presents a possible sampling bias, with some studies acknowledging limited generalizability. The higher proportion of implementation successes than failures may also indicate potential publication bias.

Stakeholder Involvement

Formal input was obtained from people with ABI (33/60, 55%), clinicians (15/60, 25%), and caregivers (14/60, 23%) in some of the reviewed studies. However, this was overwhelmingly obtained during intervention evaluation (40/42, 95%) rather than development (16/42, 38%) stages. Rietdijk et al [44] informally incorporated qualitative feedback from the users of a previous iteration of the intervention into the intervention design, in addition to formally obtaining evaluation feedback. None of the reviewed studies (0/60, 0%) involved stakeholders in coproducing research, some did not involve stakeholders at all (13/60, 22%), and a minority (5/60, 8%) provided no information.

Interventions

The most common intervention type was web-based education (10/60, 17%), including an intervention that combined education with cognitive rehabilitation [57]. Other interventions for which implementation data were available included cognitive exercises and games (8/60, 13%), Communication Partner Training (4/60, 7%), and Cognitive Behavioral Therapy (3/60, 5%). Interventions classified as *Other* (Multimedia Appendix 2; 35/60, 58%) included aphasia groups [42,43,58], emotional regulation training [54,59], and metacognitive rehabilitation [60,61]. Interventions were usually completely (26/60, 43%) or partly (21/60, 35%) individualized to the recipient's specific needs and preferences. This individualization typically occurred via the clinician (35/47, 74%) or user-selected preference (30/47, 64%). Individualization also or instead occurred through automation, such as embedded logic or artificial intelligence (15/47, 32%). This was in contrast to generic interventions that were not modified according to individual needs and preferences (7/60, 12%). Some studies provided no information about the degree of individualization (6/60, 10%). Most studies included human interaction, primarily with the clinician (39/60, 65%); however, some studies provided opportunities for peer interaction among people with ABI (14/60, 23%) and among caregivers (4/60, 7%). Some interventions involved no interaction (12/60, 20%) or only interaction with artificial intelligence (9/60, 15%).

The NASSS Framework

Domain 1: Condition

This review examined implementation evidence for interventions targeting adults with stroke (37/60, 62%), TBI (24/60, 40%), and aphasia of unspecified origin (4/60, 7%), or the formal (eg, clinicians and support workers) and informal (eg, family and partners) caregivers of this population. The psychosocial conditions under treatment included cognitive impairments, social communication difficulties, and language impairments among people with ABI and depression and caregiver burden among carers.

In each study, the nature of ABI and the psychosocial condition under treatment was almost always (56/60, 93%) *complicated* because it was "not fully characterized or understood" [17], yet not quite *complex* because participants were usually recruited in the chronic stages of injury to control for spontaneous recovery. In the remaining 7% (4/60) of studies, the condition shifted into the "unpredictable or high risk" definition of *complex* when investigators documented that participants experienced multiple neurological events [46,62,63] and responded unexpectedly to intervention during the chronic stage of injury [62,64].

Comorbidities and sociocultural factors were often (34/60, 57%) *simple* in studies because investigators set them as exclusion criteria. For example, investigators typically excluded participants with sensory or physical disabilities (including hemiparesis secondary to stroke), which would affect device use; people with intellectual disabilities; people with photosensitive epilepsy and other neurological conditions; people with mental illnesses or substance dependency; and people without carer support, device and computer proficiency, or internet access. These exclusions were acknowledged by some as a sampling bias. From among the 40% (24/60) of studies that considered or managed these comorbidities as *complicated* and *complex*, example considerations and learnings are provided in Textbox 1.

Textbox 1. Examples with citations from the reviewed studies of ways to accommodate common comorbidities.

Vision and hearing impairment [43]

- Opportunity to perform training task with researcher. Multimodal training and intervention sessions, including both auditory and visual components. All training and intervention materials used at least 14-point font, white space, and clear images. Headset microphones were used to allow users to control volume, potentially reducing the impact of background noise [43].

Cognition, memory, language, and attention [43,49,63,65,66]

- Use of desktop shortcuts to enter videoconferencing software and aphasia-friendly training materials with significant use of white space [43].
- Explicit categorization; repetition of important units of information; use of plain language and text made suitable for Australian grade-5 reading age; and following health education guidelines for people with dysphasia, such as font size and number of words per screen [46].
- Uniformity in screen design regarding backgrounds, colors, and layout. Use of accessibility and usability guidelines. Use of simple interaction methods, such as mouse clicks on big buttons, to facilitate the possibility of using the same interface for touchscreen devices [49].

Psychomotor, fine motor, and mobility [43,46,63]

- Stroke survivors suggested that the program can be developed to take into account physical ability limitations or restrictions that were due to other comorbidities [46].
- Ensuring that software or processes did not require quick responses. Appropriate positioning of required equipment in initial training session to ensure access [43].
- Participants with hemiparesis were able to make required responses with their other hand [62].

Domain 2: Technology

Implementation evidence was included from web-based interventions on a range of devices including desktop computers (35/60, 58%), tablets (7/60, 12%), desktop computers and tablets (6/60, 10%), or desktop computers and mobile devices (tablet and smartphone; 6/60, 10%). Few interventions were both tablet and smartphone apps (2/60, 3%) or a solely smartphone app (2/60, 3%). Only 3% (2/60) of studies provided no information. Half of the reviewed interventions were delivered via telehealth videoconferencing (30/60, 50%). Interventions also used a combination of text (25/60, 42%), images (19/60, 32%), audio (13/60, 22%), and video (12/60, 20%); interactive games (15/60, 25%); virtual reality (2/60, 3%); productivity tools (eg, calendar and note-taking application; 4/60, 7%); and electronic communication systems, such as instant messaging (7/60, 12%), forums or message boards (5/60, 8%), and email (11/60, 18%). Reflecting the dominance of videoconferencing, interventions were frequently clinician-led (27/60, 45%). However, almost as many (23/60, 38%) interventions were completely automated or self-guided, reflecting the substantial proportion of cognitive exercises and games represented. Other interventions (10/60, 17%) were partly automated and partly telehealth.

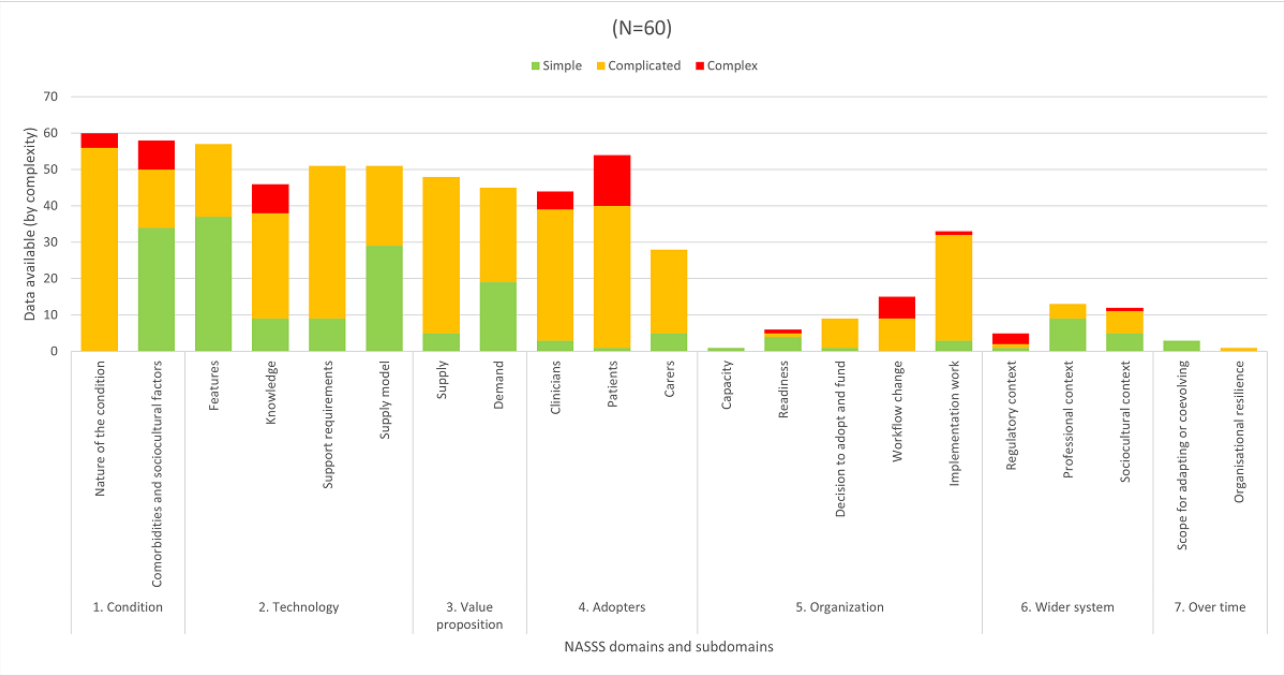
The technology was often (37/60, 62%) a *simple* off-the-shelf or preinstalled solution, including hardware provision and software installation by the researchers. Similarly, the technology supply model was often off-the-shelf or a software requiring minimal customization (29/60, 48%). Approximately one-third of technologies were *complicated* in that they were not yet fully developed or interoperable (20/60, 33%), requiring significant customization or bespoke solutions (22/60, 37%). Some studies did not provide information on supply models (9/60, 15%) or the complexity of the technology (3/60, 5%).

Although some technologies required no support or only *simple* instructions for use (9/60, 15%), most technologies (42/60, 70%) required detailed initial training and ongoing troubleshooting support. Data were mostly self-entered (39/60, 65%), but also entered by the clinician (26/60, 43%) or automatically (18/60, 30%). A minority of studies (5/60, 8%) provided no information. Most often (29/60, 48%), these data were *complicated* in that they only "partially and indirectly measured changes in the condition" [17]. In some studies (9/60, 15%), the connection between data and the condition was *simple*, "directly and transparently" measuring change [17]. In other studies (8/60, 13%), it was *complex*, when the link between data generated and changes in the condition were "unpredictable or contested" [17]. Almost a quarter of the studies (14/60, 23%) provided no information.

Domain 3: Value Proposition

With the exception of only 3 studies [64,67,68], almost all (57/60, 95%) reviewed studies included some value proposition for their intervention, including a *supply-side* case (48/60, 80%). However, almost all *supply-side* cases (43/48, 90%) were *underdeveloped* [17]. Value propositions were more likely to be *simple* in relation to *demand-side* value to end users (Figure 4). Solana et al [49] uniquely demonstrated simplicity in both *demand-side* desirability to end users and a formally calculated *supply-side* economic benefit.

Figure 4. Quantification of complexity reported in each of the domains and subdomains of the Nonadoption, Abandonment, Scale-Up, Spread, and Sustainability (NASSS) framework in the included studies, showing (1) frequent controlling of comorbidities and sociocultural factors in domain 1; (2) contentious links between clinical changes in the condition and knowledge captured by the technology in domain 2; (3) particularly complex demands of patients and clinicians to use the interventions in domain 4; and (4) limited data in domains 5 to 7, documenting the challenge of workflow changes, a need for implementation work, and some complexity in relation to technology and regulations.



Domain 4: Adopters

Adopters included in this review included 5723 adults with ABI (5424, 95% intervention recipients and 299, 5% controls), 1920 formal and informal caregivers (1729, 90% recipients and 191, 10% controls), and 50 staff (4, 8% healthy recipients, such as volunteers, who trialed the intervention; 13, 26% administrative staff in 1 study [49]; and at least 33, 66% clinicians delivering interventions, as Anderson et al [69] did not specify how many clinicians were consulted). Reviewed interventions typically targeted only the person with ABI (44/60, 73%); however, some included both an informal carer and the person with ABI (7/60, 12%), informal carers only (6/60, 10%), or either formal or informal carers together with a person with ABI (2/60, 3%) [70,71]. None of the interventions targeted only formal caregivers (ie, clinicians or support workers), but Lee et al [72] (1/60, 2%) targeted both speech-language pathology students and people with ABI.

Of the 60 studies, there were 5 (8%) studies [44,45,47,67,68] reporting *complex* requirements of clinicians in terms of expanded or altered responsibilities and scope of practice. These included new implementation responsibilities to provide constant remote monitoring and troubleshooting [45,67,68] or overall resistance to the concept of telehealth with its perceived limitations in nonverbal communication and rapport building compared to face-to-face delivery [44,47]. *Complicated* involvement (36/60, 60%) required new training, skills, and personnel. Most obviously, synchronous telehealth interventions required clinicians to learn to deliver care via the internet rather than face-to-face, including videoconferencing, instant messaging, and web-based avatars. It also required them to at least be available for technical troubleshooting. Clinicians typically delivered therapy remotely from their own home, or a private office or clinic space. There were 3 studies (3/60, 5%) in which clinician involvement was not required [12,72,73] and 16 studies (16/60, 27%) provided no information.

Expectations of people with ABI were typically (53/60, 88%) high. Most were *complicated* (39/60, 65%), with minimum expectations to log on, enter data, and converse via the internet. Almost a quarter (14/60, 23%) of the reviewed studies described more *complex* requirements, such as reflective goal setting and adjustment in response to self-monitored progress. Only 1 intervention [55] *simply* targeted carers without involving individuals with ABI.

Although falling short of *complex* demands, carers were often (23/60, 38%) assumed or required to be available whenever needed, with requirements ranging from the provision of on-demand technical support to the person with ABI to participation in more intensive programs targeting the dyad or carer themselves. A minority of studies (5/60, 8%) reported that caregiver input was not required. However, more than half of the reviewed studies (32/60, 53%) provided no information about carer involvement.

Domain 5: Organization

All except 2 (58/60, 97%) studies [72,74] examined remote delivery to homes. A few had additional availability in community health (3/60, 5%), hospitals (4/60, 7%), or university or workplace settings (3/60, 5%). Implementation scope varied from single (20/60, 33%) and multiple sites (16/60, 27%) to state-wide (9/60, 15%), national (13/60, 22%), and—rarely—international (2/60, 3%) implementation.

The organizational role in implementation was mentioned in multiple studies (38/60, 63%). However, these data were sparse (Figure 4). The sole study [69] to specify an organization's

overall capacity to innovate described a progression in service delivery from face-to-face to a hybrid telephone, self-directed, and clinic-based service, to videoconference delivery using existing facilities.

A total of 6 studies (6/60, 10%) referred to organizational readiness for technological change. Readiness was made *simple* by the use of existing infrastructure [69,75], staff experience using technology with the clinical population [41], and organizational support for the shift [76]. However, readiness was *complicated* by workflow changes [63] and became *complex* in a public health system [77], where investigators anticipated challenges "relating to treatment space, technology availability or concerns about data protection," but found the latter especially problematic due to internal regulations regarding approved information systems.

Of the reviewed studies, 15% (9/60) described an organization's adoption and funding decision, which was rarely (1/9, 11%) *simple* [69], and most commonly (8/9, 89%) *complicated*, usually by partnerships with multiple organizations. Workflow changes were described in a quarter of the reviewed studies (15/60, 25%). None of the workflow changes were *simple*, and 40% (6/15) of studies reporting workflow change described them as *complex* [41,45,52,63,68,74] due to new demands on space, time, and skill (refer to *Domain 4: Adopters* section).

Domain 6: Wider System

Investigators mentioned the *wider system* of implementation in some studies (24/60, 40%), but data were again limited (Figure 4). The wider system was mostly mentioned in relation to the position of professional bodies (13/60, 22%), followed by comments on sociocultural factors (12/60, 20%) such as internet access and technological acceptance.

In addition, a minority of studies (5/60, 8%) commented on the regulatory context. While one study made passing note of potential barriers to billing telehealth [57], the regulatory context was primarily (4/5, 80%) described in relation to security. Data security was typically (3/4, 75%) *complex* [44,63,77], with health data considerations in relation to both bespoke and off-the-shelf platforms such as Facetime (Apple Inc.) and Skype (Microsoft Corporation). The only study where this subdomain was *simple* used a legally compliant platform [49].

The only explicit mention of the political context was in relation to the widespread use of telemedicine by the American Department of Veterans Affairs [57]. Studies from several countries acknowledged government funding sources, but no other political information was available beyond state and country names.

Domain 7: Embedding and Adaptation Over Time

A minority (4/60, 7%) of studies [41,47,69,78] included data on the seventh domain. Although 3 studies [41,47,78] described a strong scope for adapting and coevolving the intervention using end user input, no information on organizational resilience was available. Inversely, although Anderson et al [69] did not provide information on adapting the intervention, theirs was the sole study to document organizational resilience in managing unforeseen complications. These included increasing bandwidth allocation when it was insufficient in the organizational network and dispatching a research assistant to resolve ad hoc technical issues on-site.

Complexity

A descriptive comparison of the records reporting successful implementation (37/45, 82%) and the those reporting failures (8/45, 18%) appears consistent with the following 3 hypotheses by Greenhalgh et al [20].

Hypothesis 1: If Most or All of the Domains Can Be Classified as Simple, an Intervention Is Likely to Be Easy to Implement and to Be Achieved on Time and Within Budget

If *most or all* were to be mathematically defined as 4 to 7 out of 7 domains, none of the reviewed studies met this definition. Therefore, it was not possible to confirm that an intervention with *most or all simple* domains was likely to succeed. Indeed, it was rare (4/60, 7%) to experience *simplicity* in >1 domain.

However, the mean number of *simple* domains and subdomains was higher for successes (mean 0.6 for domains and 3.3 for subdomains) than for *failures* (mean 0.1 for domains and 2.1 for subdomains), aligning with the possibility that more *simple* domains may increase the likelihood of implementation success.

Among failures, 88% (7/8) of the studies had no *simple* domains at all. There was only a single study (1/8, 13%) containing 1 *simple* domain. Among successes, 49% (18/37) had no *simple* domains, 43% (16/37) had 1 *simple* domain, 5% (2/37) had 2 *simple* domains, and 3% (1/37) had 3 *simple* domains. This also suggests that, even if it were true that an intervention with *most or all* simple domains is likely to succeed, the threshold for success may be much lower, and may necessarily be so, given that simplicity was so rare.

Hypothesis 2: If Many Domains Are Classified as Complicated, the Intervention Will Be Achievable but Difficult and Likely to Exceed Its Timescale and Budget

If *many domains* were to be defined as 3 to 7 domains, the hypothesis that implementation is still *achievable* with this many domains was supported; studies with as many as 6 (out of 7) *complicated* domains and 10 (out of 21) *complicated* subdomains still achieved implementation *success*. However, the hypothesis that implementation might be *difficult* could be reflected in that the degree of *complication* was similar for successes and failures at both the domain (mean 3.6 for successes and 3.4 for failures) and subdomain levels (mean 6.5 for successes and 6.6 for failures).

Hypothesis 3: If Multiple Domains Are Complex, the Chances of the Intervention Succeeding at All Are Limited

The mean number of complex domains was indeed higher among failures (mean 1.25 at both the domain and subdomain levels) than among successes (mean 0.8 at both the domain and subdomain levels). This appears to be consistent with the notion that more complex domains may hinder implementation.

Discussion

Principal Findings

In this review, we identified and appraised 60 published, peer-reviewed primary studies of implementation outcomes, strategies, or factors for web-based psychosocial interventions targeting either people with ABI, their formal or informal caregivers, or both. We narratively synthesized and quantified implementation data according to the NASSS framework to provide recommendations for future implementation. Recorded limitations of the evidence were the exclusion of individuals with a range of comorbidities and psychosocial challenges; poor methodological quality, particularly in the use of mixed methods; inconsistent use of implementation terminology; lack of theoretical underpinning; and limited data describing organizational, systemic, and long-term considerations. Our quantification of complexity across more than a decade of implementation evidence was consistent with all 3 of the following hypotheses: (1) *simplicity* facilitates implementation success; (2) successful implementation is more difficult, but still possible with more *complicated* domains; and (3) *complexity* makes implementation challenging and prone to failure [20]. These results align with recommendations to reduce complexity as much and in as many domains as possible to facilitate successful implementation [20].

Complexity and Implementation

A complexity paradigm posits that digital health implementation has many domains of complexity and that they interact. This review is the first to quantify the specific domains in which complexity has occurred and identify potential targets to improve implementation. As seen in Figure 4, most complexity in the implementation of web-based interventions for people with ABI and their caregivers was reported in the intervention demands of people with ABI and their clinicians (domain 4); the ability of the technology to measure, convey, and enable responses to health data (domain 2; Figure 4); the changes introduced to clinical workflows (domain 5); and the need to manage health data regulations (domain 6). While confirming previous findings [10,15,26,27], a complexity paradigm provides new insight that these complexities may be counterbalanced by simplifying other domains and subdomains to enable implementation success. This included simplifying domain 1 (the condition) by excluding certain comorbidities and sociocultural factors and simplifying domain 2 (the technology) by selecting *off-the-shelf* products.

The Importance of Intervention Value Propositions

In particular, a complexity paradigm highlights the relative simplicity observed in domain 3 (the value proposition) and the potential role of this simplicity in overall implementation success. Our results corroborate other digital health implementation reviews in finding that, despite their critical importance to sustainability [14], economic cases and business models for digital interventions were rarely articulated [22-24]. Financial viability remains a key challenge in digital health implementation [11], as the development and implementation of digital health interventions incur both up-front and ongoing costs. The low number of studies examining implementation beyond initially controlled studies may reflect overall sustainability challenges [11] in financially progressing past initial product development and testing [14]. Investigators seeking to improve the communication of this *supply-side* value proposition (domain 3) may benefit from both collaboration with stakeholders to identify and articulate an intervention's economic value and interdisciplinary knowledge in health economics, business, and marketing. In the absence of such an economic case, a complexity perspective highlights that demonstrations of *demand-side* value, including measures of participant satisfaction, acceptability, and feasibility, become especially important to initially establish. The reviewed evidence primarily contributed to the value proposition in this subdomain, thus simplifying domain 3.

Stakeholder Inclusion

The paradox is that *demand-side* value propositions were effectively undermined in the reviewed evidence by reductions in scope. Given the simplification of conditions (domain 1) is not possible in real-world settings, research participant exclusions based on specific comorbidities and sociocultural factors threaten the external validity of interventions [79] and reduce their potential reach due to an unrepresentative population. The exclusion of individuals with comorbidities can in fact disqualify the overwhelming majority of, and sometimes almost all, individuals with a target condition [79-81]. Such exclusions are therefore considered an increasingly untenable practice even in efficacy research, particularly given global population aging and increasing multimorbidity [80]. Continuation of this practice effectively creates a mismatch between the available evidence and clinical populations with ABI [81], the majority of which present with comorbidities [81] and psychosocial challenges [6-8,82]. Currently, this leaves clinicians and researchers ill-equipped to adequately understand how web-based psychosocial interventions may or may not be implemented with a real-world population.

Our finding of the simplification of domain 1 (the condition) corroborates reports in other systematic reviews of an unaddressed digital divide and noticeable lack of data pertaining to individuals with comorbidities in digital health implementation research more broadly [26,83]. In a recent review of digital mental health implementation [83], the most established category of digital psychosocial support [84], Barnett et al [83] discovered an absence of evidence pertaining to individuals with comorbidities and limited data on ethnic minorities. Therefore, researchers or practitioners seeking implementation data pertaining to the excluded conditions will need to seek implementation studies where such excluded conditions are the primary, rather than comorbid condition. For example, a recent systematic review of digital health implementation for individuals with psychosis or bipolar disorder [27] identified that the complexity of digital health interventions can be challenging for people with psychiatric symptoms. Given that our review identified similarly complex demands of interventions for individuals with a primary diagnosis of ABI (domain 4; Figure 4), it may be possible to see how *comorbid* psychosis and ABI diagnoses can create further complex interactions between domains 1 (the condition)

and 4 (the adopters) that were not identified in the reviewed evidence.

The underrepresentation of clinical complexities also raises ethical questions of equity [26,83] in the development of interventions for people with ABI. If no interventions are designed or tested with these excluded populations, there is a risk of perpetuating and exacerbating the disadvantages experienced by individuals already at greater risk of digital exclusion [26,83]. In the words of one of our research collaborators, Mrs Erin Elizabeth Hill, who has living experience of ABI:

> *If your aim is to help people with ABI, then you can't exclude a whole group of us, when there are more of us that have these conditions than don't. You can't say, "Oh, we considered 'some' of you." You need to be as inclusive as possible.*

Similar reviews in other populations have concluded that there is an urgent need for data pertaining to marginalized groups in digital health implementation research [26,83]. Sampling and publication biases across a body of reviewed evidence risk reducing the visibility of individuals who can be "forgotten when taking the findings of this review into consideration" [26]. Therefore, it may be helpful for researchers and clinicians to be mindful of this gap at the study design stage [79,85], not only to facilitate real-world implementation, but also to ensure that future digital health intervention design, implementation, and research does not create and perpetuate inequities [26,83].

Investigators who made proactive efforts to maximize inclusion in the reviewed studies provided valuable data about how real-world implementation may or may not be achieved. In a study that originally excluded people with ABI who had a history of falls [61], researchers recognized during recruitment that such a history was common in the population that may benefit from the intervention. They subsequently adjusted their screening criteria, based on expert panel advice, to focus on self-awareness rather than fall history, enabling previously excluded people with ABI to participate. In another study that accommodated medical complexity [63], investigators documented how readmission for subsequent strokes influenced a person with ABI's adherence to a tablet-based intervention. Although the participant eventually discontinued use, implementation data concerning dropouts and reasons for discontinuation contribute an important real-world understanding of implementation. Moreover, they enabled people with ABI who wished to participate in research and receive interventions to do so to the extent that they were able. This illustrates how, in addition to empirical questions of external validity and ethical questions of equity, population exclusions may imply a divergence between the priorities and realities of researchers and stakeholders [21], thus requiring a priori effort if it is to be overcome.

Stakeholder Collaboration

Facilitating implementation despite the real-world complexity of ABI (domain 1) presents researchers with the challenge of simplifying as many of the remaining domains (domains 2-7) as possible. Direct collaboration with stakeholders may be key to this endeavor, over and above inclusion in participant samples. In this review, population disparities were magnified by the consultation of the already unrepresentative study populations during intervention evaluation rather than development, with no studies coproduced with stakeholders. However, it is frequently recommended that stakeholders are engaged from the outset of research, rather than only in intervention evaluation [27,85-87]. To support similar efforts, we have previously published [21,34] methodological guidance on how to leverage the NASSS framework to facilitate implementation input from people with ABI, their clinicians and close others.

The need for stakeholder collaboration is further supported by our finding that the most complexity was reported in the interventional and technological requirements of people with ABI and clinicians in domain 4 (Figure 4). Simplifying an intervention's demands of people with ABI presents the largest current target to reduce complexity, with stakeholder collaboration providing opportunity to identify ways to accommodate comorbidities (domain 1; Textbox 1), co-design (domain 4), and continue to streamline (domain 7) interventions. Additionally, digital health interventions introduce complex new demands on clinicians (domain 4; Figure 4), including many tasks specific to digital health, such as remote monitoring, intervention adjustment, and providing and receiving ongoing technical support. It also introduces new space, equipment, and privacy requirements for telehealth sessions, with subsequent challenges at the organizational level (domain 5). These are critical considerations given the importance of workflow in the success or failure of digital health implementation [15]. In particular, the role of carers was underreported in this review, and only a single study [49] included the input of administrative staff who may be required to support or implement these functions, suggesting that these stakeholder groups are less visible and consulted. Additional implementation work and the staff who will likely undertake such work (domain 5; Figure 4) can also be easily overlooked due to both limited data at the organizational level (domain 5) and potential for investigators to absorb implementation work in the context of a research study. Therefore, our findings highlight the importance of recognizing health care staff adopters as stakeholders. Again, these collaborations will require significant resources and funding [88], indicating a need for researchers to upskill in the communication of value propositions to funders (domain 3) and for funders to create and invest in in the resource-intensive process of stakeholder collaboration.

Theoretical Frameworks

Finally, it is recommended that investigators use implementation frameworks, and specifically digital health implementation frameworks such as user-centered design or the NASSS framework, to underpin implementation research. User-centered design may be especially pertinent given a need to simplify domain 4 (Figure 4). The NASSS framework may also facilitate consideration across multiple domains. Our results aligned with the NASSS framework's wide-ranging inclusion of considerations in digital health implementation, with the reviewed evidence revealing challenges in all 7 domains. Our findings were consistent with other reviews applying the NASSS

framework [22-24] and those using generic implementation frameworks [26-29] in revealing a general lack of implementation evidence at the organizational level and beyond (domains 5-7; Figure 4). This may reflect the concurrent finding of limited theoretical underpinning in the reviewed studies. The limited data on organizational, systemic, and long-term aspects of implementation in this and other reviews reveal a significant gap in research evidence to date, which should be addressed to support the ecological validity and implementation of future interventions.

Study Strengths and Limitations

Reviews to date have relied on narrative syntheses when examining digital health implementation [28] and complexity [22-24]. This review was the first to both deductively analyze [25] the evidence in relation to each of the 7 domains of a complexity-based framework and quantify complexity across more than a decade of digital health implementation data against published hypotheses [20]. This investigation has enabled new understanding of the complex interrelationships between domains.

This registered review followed a published protocol [30], with deviations and rationales that are reported transparently. All extracted data, search strategies, and extraction forms are transparently appended, and results are reported according to PRISMA 2020 guidelines. To further increase replicability, data were consistently extracted and appraised using published definitions [36], taxonomies [38], and tools [37]. Although other reviews have not included information on stakeholder involvement [26] and used generic implementation frameworks [26-28], our search and synthesis were both informed by stakeholder input [21] and theoretically underpinned by an implementation framework specific to digital health [17]. Unlike reviews that focused more exclusively on scale-up [24] or implementation strategies [11], this review included and classified a wide range of data on implementation factors, outcomes, and strategies. The inclusiveness of our search thus allowed previously unreviewed implementation data to be included for the first time.

The high yield and detailed extraction from >21 theoretical subdomains presented substantial feasibility challenges, requiring pragmatic protocol deviations. Extraction was challenging due to the heterogeneity and inconsistent reporting of interventions and implementation outcomes. For instance, Pierce and Steiner [89] collectively reported satisfaction, acceptability, usability, and feasibility measures as *usability*, and Anderson et al [69] measured *satisfaction* using feasibility and acceptability measures. Therefore, standardized definitions [36] were required in the extraction form (Multimedia Appendix 2). Implementation evidence also has limited alignment with the clinical paradigm of PRISMA guidelines. Given our need for an innovative, theory-based meta-synthesis and quantification of complexity, an implementation-specific review checklist with a complexity paradigm may need to be developed as scientific understanding of both implementation and complexity grows.

As our review examined prepandemic evidence published up to June 2020, it will be possible in future to compare COVID-related and postpandemic evidence with our findings. Given that the global COVID-19 pandemic has brought domain 6 to the fore, a future update of our review that allows sufficient time for implementation effort and publication from these periods may offer a unique opportunity to meta-synthesize and directly compare data from 2 distinct epochs, allowing scrutiny of the impact of national and international shifts in this domain.

A key limitation of the reviewed evidence was that the overall quality of reviewed studies was low, with challenges for external validity due to study populations that were unrepresentative of the real-world population with ABI. The resulting potential for sampling bias may reflect pressures upon researchers to establish a *demand-side* value proposition in domain 3. The high proportion of successes may also indicate possible publication bias, which may be exacerbated by the same pressure. Another limitation was that the English-speaking research team was restricted to English publications. Finally, although the results of this descriptive analysis appear consistent with all 3 hypotheses by Greenhalgh et al [20], definitive relationships could not yet be established because:

1. The sample size of this study was not adequately powered to detect an association between complexity and implementation success or failure. In particular, the limited number of failures available for review may reflect (1) the controlled nature of clinical trials, which inherently aim to minimize complexity and (2) potential publication bias toward reporting implementation success. The intensive process of deductively coding 21 subdomains of complexity in a highly heterogeneous evidence base also reduced the feasibility of obtaining a sufficiently large sample size.
2. Although some studies reported information concerning all 7 domains of complexity, none of the studies reported data from all subdomains. Information about complexity was thus incomplete for all records.
3. In the absence of a formal definition [11], implementation success or failure of each study was defined as whether the authors succeeded in achieving their specific implementation aims, which varied in scope and type (eg, usability, cost-effectiveness, and satisfaction are distinct constructs). Studies generally also did not report planned or actual budgets and timescales as described in the hypotheses by Greenhalgh et al [20].

Given these limitations, the eventual alignment between a flexible definition of success and all 3 hypotheses [20] was unexpected. It may suggest some correlation between the various implementation measures. For example, a highly appropriate intervention may also be more acceptable to recipients, receive high satisfaction ratings, and experience increased adherence and fidelity from motivated actors. This, in turn, may improve its feasibility and cost-effectiveness. Alternatively, or perhaps concurrently, the concept of complexity may be informative for a range of implementation constructs, with further research needed to understand these relationships. For example, a multiarm RCT comparing different levels of complexity might assist our understanding of complexity within implementation science [90]. Parallel reviews of digital health implementation for other health conditions and intervention types (eg, health informatics) may also be informative; we have transparently

supplied our search terms (Multimedia Appendix 1) and extraction template (Multimedia Appendix 2) to support this endeavor.

Conclusions

This study is the first attempt to deductively analyze and quantify complexity from more than a decade of primary digital health implementation evidence. Results were consistent with recommendations to facilitate implementation success by reducing complexity in as many domains as possible. To date, simplifications appear to have been made in the first domain of the NASSS framework (the condition) to advance the value proposition of interventions (domain 3). However, this may hinder the development of equitable, real-world solutions for which implementation data and end user support are currently needed. It is recommended that intervention developers collaborate with stakeholders, including people with ABI, their close others, and clinicians, from the earliest design stages, rather than only at end-evaluation, to target real-world complexities in domains 2 to 7. Recommended future research includes parallel reviews for other populations and intervention types, multiarm RCTs to test the role of complexity in digital health implementation, and reviews of evidence obtained during or after the COVID-19 pandemic to understand the impact of the wider context of digital health implementation (domain 6).

Acknowledgments

MM would like to thank Mrs Erin Elizabeth Hill for her expert advice based on her living experience of acquired brain injury, Dr Mikaela Jorgensen for her advice on statistical analysis, and Ms Rebecca Dale for her advice on search strategy. MM is supported by an Australian National Health and Medical Research Council Postgraduate (PhD) Scholarship Grant (GNT1191284) and Australian Research Training Program Scholarship. LT is supported by an Australian National Health and Medical Research Council Senior Research Fellowship. RR is supported by funding from icare NSW. The Australian National Health and Medical Research Council and icare NSW were not involved in the conception, design, or conduct of the review.

Data Availability

The search strategy is available as Multimedia Appendix 1, data extraction form is included as Multimedia Appendix 2, and data extracted from included studies and used for analyses are provided as Multimedia Appendix 3.

Authors' Contributions

MM, EP, LT, RR, and MB conceptualized the study. All authors contributed to study design. MM, MB, RR, EP, and DD independently screened records at both the abstract and full-text levels. MM completed extraction, and MB, DD, and RR checked extraction. All authors participated in consensus discussions. MM completed the analysis and drafted the manuscript. All authors critically revised the manuscript and approved the final version for submission.

Conflicts of Interest

LT is the director of speechBITE. EP and MB are board members of speechBITE.

Multimedia Appendix 1
Search strategy.
[PDF File (Adobe PDF File), 240 KB - jmir_v24i7e38100_app1.pdf]

Multimedia Appendix 2
REDCap (Research Electronic Data Capture) extraction form.
[PDF File (Adobe PDF File), 106 KB - jmir_v24i7e38100_app2.pdf]

Multimedia Appendix 3
Extracted data and analysis of complexity.
[XLSX File (Microsoft Excel File), 508 KB - jmir_v24i7e38100_app3.xlsx]

References

1. Cieza A, Causey K, Kamenov K, Hanson SW, Chatterji S, Vos T. Global estimates of the need for rehabilitation based on the Global Burden of Disease study 2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020 Dec 19;396(10267):2006-2017 [FREE Full text] [doi: 10.1016/S0140-6736(20)32340-0] [Medline: 33275908]
2. GBD 2016 Neurology Collaborators. Global, regional, and national burden of neurological disorders, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol 2019 May;18(5):459-480 [FREE Full text] [doi: 10.1016/S1474-4422(18)30499-X] [Medline: 30879893]

3. Finch E, French A, Ou RJ, Fleming J. Participation in communication activities following traumatic brain injury: a time use diary study. Brain Inj 2016 Jun;30(7):883-890. [doi: 10.3109/02699052.2016.1146959] [Medline: 27029816]
4. Hilari K, Northcott S. "Struggling to stay connected": comparing the social relationships of healthy older people and people with stroke and aphasia. Aphasiology 2016 Aug;31(6):674-687. [doi: 10.1080/02687038.2016.1218436]
5. Mitchell AJ, Sheth B, Gill J, Yadegarfar M, Stubbs B, Yadegarfar M, et al. Prevalence and predictors of post-stroke mood disorders: a meta-analysis and meta-regression of depression, anxiety and adjustment disorder. Gen Hosp Psychiatry 2017 Jul;47:48-60. [doi: 10.1016/j.genhosppsych.2017.04.001] [Medline: 28807138]
6. Prisnie JC, Sajobi TT, Wang M, Patten SB, Fiest KM, Bulloch AG, et al. Effects of depression and anxiety on quality of life in five common neurological disorders. Gen Hosp Psychiatry 2018 May;52:58-63. [doi: 10.1016/j.genhosppsych.2018.03.009] [Medline: 29684713]
7. Douglas JM, Bracy CA, Snow PC. Return to work and social communication ability following severe traumatic brain injury. J Speech Lang Hear Res 2016 Jun 01;59(3):511-520. [doi: 10.1044/2015_JSLHR-L-15-0025] [Medline: 27124205]
8. Langhammer B, Sunnerhagen KS, Sällström S, Becker F, Stanghelle JK. Return to work after specialized rehabilitation-an explorative longitudinal study in a cohort of severely disabled persons with stroke in seven countries. Brain Behav 2018 Jul;8(8):e01055 [FREE Full text] [doi: 10.1002/brb3.1055] [Medline: 30022609]
9. Buchan H. Gaps between best evidence and practice: causes for concern. Med J Aust 2004 Mar 15;180(S6):S48-S49. [doi: 10.5694/j.1326-5377.2004.tb05944.x] [Medline: 15012579]
10. Ahmed B, Dannhauser T, Philip N. A systematic review of reviews to identify key research opportunities within the field of eHealth implementation. J Telemed Telecare 2019 Jun;25(5):276-285. [doi: 10.1177/1357633X18768601] [Medline: 29703127]
11. Varsi C, Solberg Nes L, 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;21(9):e14255 [FREE Full text] [doi: 10.2196/14255] [Medline: 31573934]
12. Munsell M, De Oliveira E, Saxena S, Godlove J, Kiran S. Closing the digital divide in speech, language, and cognitive therapy: cohort study of the factors associated with technology usage for rehabilitation. J Med Internet Res 2020 Feb;22(2):e16286 [FREE Full text] [doi: 10.2196/16286] [Medline: 32044752]
13. Titov N, Dear BF, Nielssen O, Wootton B, Kayrouz R, Karin E, et al. User characteristics and outcomes from a national digital mental health service: an observational study of registrants of the Australian MindSpot Clinic. Lancet Digit Health 2020 Nov;2(11):e582-e593 [FREE Full text] [doi: 10.1016/S2589-7500(20)30224-7] [Medline: 33103097]
14. Christie HL, Martin JL, Connor J, Tange HJ, Verhey FR, de Vugt ME, et al. eHealth interventions to support caregivers of people with dementia may be proven effective, but are they implementation-ready? Internet Interv 2019 Dec;18:100260 [FREE Full text] [doi: 10.1016/j.invent.2019.100260] [Medline: 31890613]
15. 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;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
16. Papoutsi C, Wherton J, Shaw S, Morrison C, Greenhalgh T. Putting the social back into sociotechnical: case studies of co-design in digital health. J Am Med Inform Assoc 2021 Feb;28(2):284-293 [FREE Full text] [doi: 10.1093/jamia/ocaa197] [Medline: 33043359]
17. 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]
18. Cohn S, Clinch M, Bunn C, Stronge P. Entangled complexity: why complex interventions are just not complicated enough. J Health Serv Res Policy 2013 Jan;18(1):40-43. [doi: 10.1258/jhsrp.2012.012036] [Medline: 23393041]
19. Papoutsi C, Shaw J, Paparini S, Shaw S. We need to talk about complexity in health research: findings from a focused ethnography. Qual Health Res 2021 Jan;31(2):338-348 [FREE Full text] [doi: 10.1177/1049732320968779] [Medline: 33155510]
20. Greenhalgh T, Abimbola S. The NASSS framework - a synthesis of multiple theories of technology implementation. Stud Health Technol Inform 2019 Jul;263:193-204. [doi: 10.3233/SHTI190123] [Medline: 31411163]
21. Miao M, Power E, Rietdijk R, Debono D, Brunner M, Salomon A, et al. Coproducing knowledge of the implementation of complex digital health interventions for adults with acquired brain injury and their communication partners: protocol for a mixed methods study. JMIR Res Protoc 2022 Jan;11(1):e35080 [FREE Full text] [doi: 10.2196/35080] [Medline: 35006082]
22. Bastoni S, Wrede C, da Silva MC, Sanderman R, Gaggioli A, Braakman-Jansen A, et al. Factors influencing implementation of eHealth technologies to support informal dementia care: umbrella review. JMIR Aging 2021 Oct;4(4):e30841 [FREE Full text] [doi: 10.2196/30841] [Medline: 34623314]
23. Renyi M, Lindwedel-Reime U, Blattert L, Teuteberg F, Kunze C. Collaboration applications for mixed home care - a systematic review of evaluations and outcomes. Int J Technol Assess Health Care 2020 Aug;36(4):395-403 [FREE Full text] [doi: 10.1017/S0266462320000458] [Medline: 32686628]

24. 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;23(1):e23775 [FREE Full text] [doi: 10.2196/23775] [Medline: 33434141]
25. Kyngäs H, Kaakinen P. Deductive content analysis. In: Kyngäs H, Mikkonen K, Kääriäinen M, editors. The Application of Content Analysis in Nursing Science Research. Cham, Switzerland: Springer International Publishing; 2020:23-30.
26. Appleton R, Williams J, Vera San Juan N, Needle JJ, Schlief M, Jordan H, et al. Implementation, adoption, and perceptions of telemental health during the COVID-19 pandemic: systematic review. J Med Internet Res 2021 Dec;23(12):e31746 [FREE Full text] [doi: 10.2196/31746] [Medline: 34709179]
27. Aref-Adib G, McCloud T, Ross J, O'Hanlon P, Appleton V, Rowe S, et al. Factors affecting implementation of digital health interventions for people with psychosis or bipolar disorder, and their family and friends: a systematic review. Lancet Psychiatry 2019 Mar;6(3):257-266 [FREE Full text] [doi: 10.1016/S2215-0366(18)30302-X] [Medline: 30522979]
28. 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;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
29. Christie HL, Bartels SL, Boots LM, Tange HJ, Verhey FR, de Vugt ME. A systematic review on the implementation of eHealth interventions for informal caregivers of people with dementia. Internet Interv 2018 Sep;13:51-59 [FREE Full text] [doi: 10.1016/j.invent.2018.07.002] [Medline: 30206519]
30. Miao M, Power E, Rietdijk R, Brunner M, Togher L. Implementation of online psychosocial interventions for people with neurological conditions and their caregivers: a systematic review protocol. Digit Health 2021 Sep;7:20552076211035988 [FREE Full text] [doi: 10.1177/20552076211035988] [Medline: 34567610]
31. Miao M, Power E, Rietdijk R, Brunner M, Togher L. Implementation of online psychosocial interventions for people with neurological conditions and their caregivers: a systematic review protocol. PROSPERO International prospective register of systematic reviews. 2020 Jul. URL: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020186387 [accessed 2020-09-24]
32. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. PLoS Med 2021 Mar;18(3):e1003583 [FREE Full text] [doi: 10.1371/journal.pmed.1003583] [Medline: 33780438]
33. Covidence. Melbourne, Australia: Veritas Health Innovation; 2022. URL: https://www.covidence.org/home [accessed 2020-05-14]
34. Miao M, Power E, Rietdijk R, Brunner M, Debono D, Togher L. A Web-based service delivery model for communication training after brain injury: protocol for a mixed methods, prospective, hybrid type 2 implementation-effectiveness study. JMIR Res Protoc 2021 Dec;10(12):e31995 [FREE Full text] [doi: 10.2196/31995] [Medline: 34889770]
35. Connor BB, Shaw CA. Case study series using brain-training games to treat attention and memory following brain injury. J Pain Manag 2016;9(3):217-226 [FREE Full text]
36. 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]
37. 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.e1 [FREE Full text] [doi: 10.1016/j.jclinepi.2019.03.008] [Medline: 30905698]
38. Bewick BM, Ondersma SJ, Høybye MT, Blakstad O, Blankers M, Brendryen H, et al. Key intervention characteristics in e-Health: steps towards standardized communication. Int J Behav Med 2017 Apr;24(5):659-664 [FREE Full text] [doi: 10.1007/s12529-016-9630-3] [Medline: 28405917]
39. 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]
40. Cikajlo I, Cizman Staba U, Vrhovac S, Larkin F, Roddy M. A cloud-based virtual reality app for a novel telemindfulness service: rationale, design and feasibility evaluation. JMIR Res Protoc 2017 Jun;6(6):e108 [FREE Full text] [doi: 10.2196/resprot.6849] [Medline: 28583904]
41. Steele RD, Baird A, McCall D, Haynes L. Combining teletherapy and on-line language exercises in the treatment of chronic aphasia: an outcome study. Int J Telerehabil 2014 Jan;6(2):3-20 [FREE Full text] [doi: 10.5195/ijt.2014.6157] [Medline: 25945225]
42. Pitt R, Hill AJ, Theodoros D, Russell T. "I definitely think it's a feasible and worthwhile option": perspectives of speech-language pathologists providing online aphasia group therapy. Aphasiology 2018 Jun;32(9):1031-1053 [FREE Full text] [doi: 10.1080/02687038.2018.1482403]
43. Pitt R, Theodoros D, Hill AJ, Russell T. The development and feasibility of an online aphasia group intervention and networking program - TeleGAIN. Int J Speech Lang Pathol 2019 Sep;21(1):23-36 [FREE Full text] [doi: 10.1080/17549507.2017.1369567] [Medline: 28868932]

44. Rietdijk R, Power E, Brunner M, Togher L. A single case experimental design study on improving social communication skills after traumatic brain injury using communication partner telehealth training. Brain Inj 2019;33(1):94-104. [doi: 10.1080/02699052.2018.1531313] [Medline: 30325220]
45. Oh H, Park J, Seo W. Development of an evidence-based intervention program for patients with mild cognitive impairment after stroke: a home-based, online cognitive rehabilitation program. Top Geriatr Rehabil 2017 Apr;33(2):140-151 [FREE Full text] [doi: 10.1097/tgr.0000000000000149]
46. Denham AM, Guillaumier A, McCrabb S, Turner A, Baker AL, Spratt NJ, et al. Development of an online secondary prevention programme for stroke survivors: prevent 2nd stroke. BMJ Innov 2019 Mar;5(1):35-42 [FREE Full text] [doi: 10.1136/bmjinnov-2017-000257]
47. Simic T, Leonard C, Laird L, Cupit J, Höbler F, Rochon E. A usability study of Internet-based therapy for naming deficits in aphasia. Am J Speech Lang Pathol 2016 Nov;25(4):642-653 [FREE Full text] [doi: 10.1044/2016_AJSLP-15-0030] [Medline: 27784033]
48. Caunca MR, Simonetto M, Hartley G, Wright CB, Czaja SJ. Design and usability testing of the stroke caregiver support system: a mobile-friendly website to reduce stroke caregiver burden. Rehabil Nurs 2020 May;45(3):166-177 [FREE Full text] [doi: 10.1097/RNJ.0000000000000196] [Medline: 30418319]
49. Solana J, Cáceres C, García-Molina A, Opisso E, Roig T, Tormos JM, et al. Improving brain injury cognitive rehabilitation by personalized telerehabilitation services: Guttmann neuropersonal trainer. IEEE J Biomed Health Inform 2015 Jan;19(1):124-131 [FREE Full text] [doi: 10.1109/JBHI.2014.2354537] [Medline: 25204001]
50. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs. Med Care 2012 Mar;50(3):217-226 [FREE Full text] [doi: 10.1097/MLR.0b013e3182408812] [Medline: 22310560]
51. Ng EM, Polatajko HJ, Marziali E, Hunt A, Dawson DR. Telerehabilitation for addressing executive dysfunction after traumatic brain injury. Brain Inj 2013 May;27(5):548-564 [FREE Full text] [doi: 10.3109/02699052.2013.766927] [Medline: 23472964]
52. Raina KD, Morse JQ, Chisholm D, Leibold ML, Shen J, Whyte E. Feasibility of a cognitive behavioral intervention to manage fatigue in individuals with traumatic brain injury: a pilot study. J Head Trauma Rehabil 2016 Sep;31(5):E41-E49 [FREE Full text] [doi: 10.1097/HTR.0000000000000196] [Medline: 26580691]
53. Sharma B, Tomaszczyk JC, Dawson D, Turner GR, Colella B, Green RE. Feasibility of online self-administered cognitive training in moderate-severe brain injury. Disabil Rehabil 2017 Jul;39(14):1380-1390 [FREE Full text] [doi: 10.1080/09638288.2016.1195453] [Medline: 27414703]
54. Tsaousides T, D'Antonio E, Varbanova V, Spielman L. Delivering group treatment via videoconference to individuals with traumatic brain injury: a feasibility study. Neuropsychol Rehabil 2014 May;24(5):784-803 [FREE Full text] [doi: 10.1080/09602011.2014.907186] [Medline: 24810148]
55. Damianakis T, Tough A, Marziali E, Dawson DR. Therapy online: a Web-based video support group for family caregivers of survivors with traumatic brain injury. J Head Trauma Rehabil 2016 Jul;31(4):E12-E20 [FREE Full text] [doi: 10.1097/HTR.0000000000000178] [Medline: 26291634]
56. Withiel TD, Sharp VL, Wong D, Ponsford JL, Warren N, Stolwyk RJ. Understanding the experience of compensatory and restorative memory rehabilitation: a qualitative study of stroke survivors. Neuropsychol Rehabil 2020 Apr;30(3):503-522. [doi: 10.1080/09602011.2018.1479275] [Medline: 29886809]
57. Riegler L, Wade SL, Narad M, Sarver L. POWER-A telehealth rehabilitation program for veterans: feasibility and preliminary efficacy. Perspect ASHA SIGs 2017 Jan;2(18):3-14 [FREE Full text] [doi: 10.1044/persp2.sig18.3]
58. Walker JP, Price K, Watson J. Promoting social connections in a synchronous telepractice, aphasia communication group. Perspect ASHA SIGs 2018 Jan;3(18):32-42 [FREE Full text] [doi: 10.1044/persp3.sig18.32]
59. Tsaousides T, Spielman L, Kajankova M, Guetta G, Gordon W, Dams-O Connor K. Improving emotion regulation following Web-based group intervention for individuals with traumatic brain injury. J Head Trauma Rehabil 2017 Sep;32(5):354-365 [FREE Full text] [doi: 10.1097/HTR.0000000000000345] [Medline: 28891909]
60. Beit Yosef A, Jacobs JM, Shenkar S, Shames J, Schwartz I, Doryon Y, et al. Activity performance, participation, and quality of life among adults in the chronic stage after acquired brain injury-the feasibility of an occupation-based telerehabilitation intervention. Front Neurol 2019 Dec;10:1247 [FREE Full text] [doi: 10.3389/fneur.2019.01247] [Medline: 31866924]
61. Kringle EA, Setiawan IM, Golias K, Parmanto B, Skidmore ER. Feasibility of an iterative rehabilitation intervention for stroke delivered remotely using mobile health technology. Disabil Rehabil Assist Technol 2019 Jun;15(8):908-916 [FREE Full text] [doi: 10.1080/17483107.2019.1629113] [Medline: 31216917]
62. Peers PV, Astle DE, Duncan J, Murphy FC, Hampshire A, Das T, et al. Dissociable effects of attention vs working memory training on cognitive performance and everyday functioning following fronto-parietal strokes. Neuropsychol Rehabil 2020 Jul;30(6):1092-1114 [FREE Full text] [doi: 10.1080/09602011.2018.1554534] [Medline: 30569816]
63. Pugliese M, Ramsay T, Shamloul R, Mallet K, Zakutney L, Corbett D, et al. RecoverNow: a mobile tablet-based therapy platform for early stroke rehabilitation. PLoS One 2019 Jan;14(1):e0210725 [FREE Full text] [doi: 10.1371/journal.pone.0210725] [Medline: 30682076]

64. Withiel TD, Wong D, Ponsford JL, Cadilhac DA, Stolwyk RJ. Feasibility and effectiveness of computerised cognitive training for memory dysfunction following stroke: a series of single case studies. Neuropsychol Rehabil 2020 Jun;30(5):829-852. [doi: 10.1080/09602011.2018.1503083] [Medline: 30058468]
65. Getz H, Snider S, Brennan D, Friedman R. Successful remote delivery of a treatment for phonological alexia via telerehab. Neuropsychol Rehabil 2016 Aug;26(4):584-609 [FREE Full text] [doi: 10.1080/09602011.2015.1048254] [Medline: 26018197]
66. Topolovec-Vranic J, Cullen N, Michalak A, Ouchterlony D, Bhalerao S, Masanic C, et al. Evaluation of an online cognitive behavioural therapy program by patients with traumatic brain injury and depression. Brain Inj 2010;24(5):762-772. [doi: 10.3109/02699051003709599] [Medline: 20370383]
67. Zannino GD, Zabberoni S, Murolo R, Caltagirone C, Carlesimo GA. Picture and spoken word presentation in repetition training for anomia: does stimulus order matter? Evidence obtained from 12 individuals with chronic aphasia using a computer-based telemedicine approach. Aphasiology 2020;34(3):275-299 [FREE Full text] [doi: 10.1080/02687038.2018.1563876]
68. Wentink MM, Meesters J, Berger MA, de Kloet AJ, Stevens E, Band GP, et al. Adherence of stroke patients with an online brain training program: the role of health professionals' support. Top Stroke Rehabil 2018 Jul;25(5):359-365. [doi: 10.1080/10749357.2018.1459362] [Medline: 29663857]
69. Anderson J, Godwin KM, Petersen NJ, Willson P, Kent TA. A pilot test of videoconferencing to improve access to a stroke risk-reduction programme for Veterans. J Telemed Telecare 2013 May;19(3):153-159 [FREE Full text] [doi: 10.1177/1357633X13479703] [Medline: 23625913]
70. Rietdijk R, Power E, Attard M, Heard R, Togher L. Improved conversation outcomes after social communication skills training for people with traumatic brain injury and their communication partners: a clinical trial investigating in-person and telehealth delivery. J Speech Lang Hear Res 2020 Feb;63(2):615-632 [FREE Full text] [doi: 10.1044/2019_jslhr-19-00076]
71. Rietdijk R, Power E, Attard M, Togher L. Acceptability of telehealth-delivered rehabilitation: experiences and perspectives of people with traumatic brain injury and their carers. J Telemed Telecare 2022 Feb;28(2):122-134. [doi: 10.1177/1357633X20923824] [Medline: 32460583]
72. Lee JJ, Finch E, Rose T. Exploring the outcomes and perceptions of people with aphasia who conversed with speech pathology students via telepractice: a pilot study. Speech Lang Hear 2019 Dec;23(2):110-120 [FREE Full text] [doi: 10.1080/2050571x.2019.1702241]
73. McLaughlin KA, Glang A, Beaver SV, Gau JM, Keen S. Web-based training in family advocacy. J Head Trauma Rehabil 2013;28(5):341-348 [FREE Full text] [doi: 10.1097/HTR.0b013e31824e1d43] [Medline: 22688211]
74. Sander AM, Clark AN, Atchison TB, Rueda M. A Web-based videoconferencing approach to training caregivers in rural areas to compensate for problems related to traumatic brain injury. J Head Trauma Rehabil 2009;24(4):248-261. [doi: 10.1097/HTR.0b013e3181ad593a] [Medline: 19625864]
75. McDonald S, Trimmer E, Newby J, Grant S, Gertler P, Simpson GK. Providing on-line support to families of people with brain injury and challenging behaviour: a feasibility study. Neuropsychol Rehabil 2021 Apr;31(3):392-413. [doi: 10.1080/09602011.2019.1696846] [Medline: 31778091]
76. Quaco C. Using a telehealth service delivery approach to working with an undergraduate student with a traumatic brain injury: a case study. Work 2017 Sep 14;58(1):17-21. [doi: 10.3233/WOR-172600] [Medline: 28922179]
77. Woolf C, Caute A, Haigh Z, Galliers J, Wilson S, Kessie A, et al. A comparison of remote therapy, face to face therapy and an attention control intervention for people with aphasia: a quasi-randomised controlled feasibility study. Clin Rehabil 2016 Apr;30(4):359-373. [doi: 10.1177/0269215515582074] [Medline: 25911523]
78. Hill AJ, Breslin HM. Refining an asynchronous telerehabilitation platform for speech-language pathology: engaging end-users in the process. Front Hum Neurosci 2016;10:640 [FREE Full text] [doi: 10.3389/fnhum.2016.00640] [Medline: 28066211]
79. He Z, Tang X, Yang X, Guo Y, George TJ, Charness N, et al. Clinical trial generalizability assessment in the big data era: a review. Clin Transl Sci 2020 Jul;13(4):675-684 [FREE Full text] [doi: 10.1111/cts.12764] [Medline: 32058639]
80. He J, Morales DR, Guthrie B. Exclusion rates in randomized controlled trials of treatments for physical conditions: a systematic review. Trials 2020 Feb 26;21(1):228 [FREE Full text] [doi: 10.1186/s13063-020-4139-0] [Medline: 32102686]
81. Jackson HM, Troeung L, Martini A. Prevalence, patterns, and predictors of multimorbidity in adults with acquired brain injury at admission to staged community-based rehabilitation. Arch Rehabil Res Clin Transl 2020 Dec;2(4):100089 [FREE Full text] [doi: 10.1016/j.arrct.2020.100089] [Medline: 33543112]
82. Young JT, Hughes N. Traumatic brain injury and homelessness: from prevalence to prevention. Lancet Public Health 2020 Jan;5(1):e4-e5 [FREE Full text] [doi: 10.1016/S2468-2667(19)30225-7] [Medline: 31806488]
83. Barnett P, Goulding L, Casetta C, Jordan H, Sheridan-Rains L, Steare T, et al. Implementation of telemental health services before COVID-19: rapid umbrella review of systematic reviews. J Med Internet Res 2021 Jul 20;23(7):e26492 [FREE Full text] [doi: 10.2196/26492] [Medline: 34061758]
84. Andersson G. Internet interventions: past, present and future. Internet Interv 2018 Jun;12:181-188 [FREE Full text] [doi: 10.1016/j.invent.2018.03.008] [Medline: 30135782]

85. Landes SJ, McBain SA, Curran GM. Reprint of: an introduction to effectiveness-implementation hybrid designs. Psychiatry Res 2020 Jan;283:112630 [FREE Full text] [doi: 10.1016/j.psychres.2019.112630] [Medline: 31722790]
86. Statement on consumer and community involvement in health and medical research. National Health and Medical Research Council. 2016 Sep. URL: https://www.nhmrc.gov.au/about-us/publications/statement-consumer-and-community-involvement-health-and-medical-research [accessed 2021-10-24]
87. UNESCO Recommendation on Open Science. United Nations Educational, Scientific and Cultural Organization. 2021. URL: https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en [accessed 2021-12-23]
88. Mulvale G, Moll S, Miatello A, Robert G, Larkin M, Palmer VJ, et al. Codesigning health and other public services with vulnerable and disadvantaged populations: insights from an international collaboration. Health Expect 2019 Jun;22(3):284-297 [FREE Full text] [doi: 10.1111/hex.12864] [Medline: 30604580]
89. Pierce LL, Steiner V. Usage and design evaluation by family caregivers of a stroke intervention web site. J Neurosci Nurs 2013 Oct;45(5):254-261 [FREE Full text] [doi: 10.1097/JNN.0b013e31829dba61] [Medline: 24025464]
90. Wolfenden L, Foy R, Presseau J, Grimshaw JM, Ivers NM, Powell BJ, et al. Designing and undertaking randomised implementation trials: guide for researchers. BMJ 2021 Jan 18;372:m3721 [FREE Full text] [doi: 10.1136/bmj.m3721] [Medline: 33461967]

Abbreviations

ABI: acquired brain injury
MMAT: Mixed Methods Appraisal Tool
NASSS: Nonadoption, Abandonment, Scale-Up, Spread, and Sustainability
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture
TBI: traumatic brain injury

Edited by R Kukafka; submitted 19.03.22; peer-reviewed by A Bamgboje-Ayodele, A Miatello; comments to author 05.05.22; revised version received 16.05.22; accepted 15.06.22; published 26.07.22.

Please cite as:
Miao M, Rietdijk R, Brunner M, Debono D, Togher L, Power E
Implementation of Web-Based Psychosocial Interventions for Adults With Acquired Brain Injury and Their Caregivers: Systematic Review
J Med Internet Res 2022;24(7):e38100
URL: https://www.jmir.org/2022/7/e38100
doi:10.2196/38100
PMID:35881432

Review

The Role of Online Support Groups in Helping Individuals Affected by HIV and AIDS: Scoping Review of the Literature

Neil S Coulson[1*], MA, MSc, PhD; Heather Buchanan[1*], BA, MSc, PhD

School of Medicine, Faculty of Medicine and Health Sciences, University of Nottingham, Nottingham, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Neil S Coulson, MA, MSc, PhD
School of Medicine
Faculty of Medicine and Health Sciences
University of Nottingham
Clinical Sciences Building 2
Nottingham City Hospital
Nottingham, NG5 1PB
United Kingdom
Phone: 44 0115 84 66642
Email: neil.coulson@nottingham.ac.uk

Abstract

Background: Online support groups provide opportunities for individuals affected by HIV and AIDS to seek information, advice, and support from peers. However, whether and how engagement with online support groups helps individuals affected by HIV and AIDS remains unclear, as does the nature of the evidence on this topic.

Objective: This scoping review sought to explore whether engagement with HIV and AIDS–related online support groups benefits members in terms of psychosocial well-being and illness management, whether members experienced any negative aspects of these groups, and what types of social support are exchanged within HIV and AIDS–related online support groups.

Methods: A scoping review of English-language articles (including both qualitative and quantitative studies) was undertaken using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. The databases searched included MEDLINE, PubMed, EMBASE, CINAHL, PsycINFO, CENTRAL (Cochrane Register of Controlled Trials), and Scopus. Key findings were synthesized using a narrative and thematic approach.

Results: A total of 22 papers met the inclusion criteria from an initial pool of 3332 abstracts. These papers included 23% (5/22) quantitative studies, 9% (2/22) mixed methods studies, and 68% (15/22) qualitative studies published between 2007 and 2019. Cross-sectional evidence suggests that engagement with HIV and AIDS–related online support groups is empowering for members and may lead to a range of psychosocial benefits. Furthermore, qualitative evidence suggests that these groups provide an opportunity to connect with similar people and share experiences. This can help improve self-worth, reduce stigma, facilitate improved illness management, and gain greater confidence when interacting with health professionals. However, online support groups are not without their limitations as qualitative evidence suggests that users may encounter examples of interpersonal conflict between members as well as be exposed to challenging content. Finally, HIV and AIDS–related online support groups are avenues through which individuals can solicit support, most commonly informational or emotional.

Conclusions: HIV and AIDS–related online support groups may have some benefits for members, particularly in terms of providing social support. There is a need for a systematic review of this literature that includes an assessment of the methodological quality of the available evidence.

(J Med Internet Res 2022;24(7):e27648) doi:10.2196/27648

KEYWORDS

AIDS; HIV; online support groups; internet; peer support; social support; synthesis; systematic review

Introduction

Background

According to the Joint United Nations Programme on HIV and AIDS, it is estimated that approximately 38 million people are living with HIV and AIDS worldwide [1]. Although there is currently no cure, it is possible to suppress the virus to levels that are undetectable using antiretroviral drugs. Nevertheless, a new diagnosis of HIV brings with it several challenges [2,3]. Individuals receiving a positive diagnosis of HIV will likely face a lifetime of medical treatment to combat the biomedical repercussions of the disease. Individuals also face many psychological challenges because of living with a long-term, highly stigmatized condition [4,5], which may lead to uncertainty about the future. Individuals living with HIV and AIDS may experience social isolation [6], fear of prejudice [7], and loneliness [8] as they learn to adjust to their diagnosis and a lifetime of daily multiple medications [9]. If individuals have problems adjusting to living with HIV and AIDS, clinical depression, anxiety, stress, and poor coping are common [10]. Of particular concern is the prevalence of depression and suicide among people living with HIV and AIDS [11-13].

Evidence suggests that individuals living with HIV and AIDS who are satisfied with the level of social support they receive are more likely to adjust positively, cope better, and experience a slower progression of HIV-related symptoms [14]. Furthermore, research has shown that social support plays a pivotal role in managing the stress associated with HIV and leads to better psychological and physical health outcomes among individuals living with the disease [15,16].

As global access to the internet continues to increase [17], recent technological advances have led to the development of diverse forms of electronic communication, which in turn have supported participation, collaboration, and information sharing between users. An example that illustrates the potential for users to interact with peers on the web is through the medium of online support groups. Such groups allow individuals to come together to share experiences, provide mutual support, and ask questions. Online support groups can be underpinned by different platforms such as discussion forums, chat rooms, social networking sites (eg, Facebook), blogs, microblogs (eg, Twitter), and virtual reality environments. Although such platforms may offer synchronous interaction (ie, live and in real time), most online support groups are asynchronous, where interaction and the exchange of user-generated content takes place over time (ie, hours, days, weeks, or months), and are predominantly text based.

In recent years, there has been an exponential increase in the number of online support groups that have been established to help those affected by long-term conditions, including HIV and AIDS. Similarly, the number of internet users accessing online support groups continues to increase, with recent estimates suggesting that between 7% and 28% of adults have accessed one [18,19].

Evidence suggests that there may be a range of factors that can lead individuals to engage with asynchronous text-based online support groups. In a review of the literature, Wright [20] described 4 broad factors evident within the studies. First, the convenience of computer-mediated communication may be attractive to individuals. For example, an asynchronous text-based online support group is potentially available 24 hours per day, 7 days per week and can be accessed whenever it is needed [21]. This flexibility in access permits individuals to seek support at times and places that are convenient to them and may be helpful to those with family, educational, or work commitments [22-24]. Second, individuals may have limited access to adequate social support within traditional social networks. This may be because people in an individual's social network have little or no experience or understanding of their condition. Indeed, their condition may not be well understood by health professionals, or it may be rare. Third, individuals may be living with a condition that is stigmatized and, therefore, online support groups may be regarded as a safe environment in which they can discuss personal or sensitive issues [25,26]. Fourth, individuals have reported the value of being able to interact with others who are similar and credible [27]. Taken together, it is evident that online support groups may be relevant and potentially beneficial for individuals living with HIV and AIDS.

However, it should be noted that online support groups are not without their limitations. For example, the asynchronous text-based nature of most online support groups means that social cues such as facial expressions, tone of voice, and body language are not available, and this may cause challenges for users [28]. In addition, the absence of physical proximity restricts the expression of physical affection (eg, hugs), and users may feel isolated and alone in their real lives after logging off [24]. There may also be delays in responses being posted to the group, and this may negatively affect the user experience and satisfaction with online support groups [21]. Concerns have also been expressed regarding both the quantity and quality of posts and information [29,30]. Finally, there may exist a dominance of negative content as users are more likely to post messages during times when their symptoms are especially problematic, and this may cause additional anxiety and concern for those reading these posts [24].

Rationale for the Study

The internet affords new opportunities to support those living with a stigmatized condition such as HIV and AIDS. Online support groups provide a convenient, anonymous, and increasingly popular way to reach out to similar people for information, advice, and support. In addition, increasing attention has been given to online support groups by policy makers [31]. However, there has been no attempt to review the evidence on the possible benefits (or limitations) of HIV and AIDS–related online support groups as well as the types of social support that may be exchanged between users. In this scoping review, both the quantitative and qualitative literature will be identified, described, and synthesized.

Aim and Review Questions

The primary aim of this scoping review was to describe the literature on the utility of online peer support groups for individuals affected by HIV and AIDS, and 3 key questions

informed our review. We did not make any assumptions about whether quantitative or qualitative studies would ultimately be used to address each question. Rather, our scoping review set out to establish what types of evidence existed that could help address each of the research questions. Our questions were as follows: (1) Does engagement with online peer support groups improve psychosocial well-being and illness management among those living with HIV and AIDS? (2) Are there any negative aspects of online support groups experienced by individuals living with or affected by HIV and AIDS? If so, what are these? (3) What types of social support are exchanged within online support groups for individuals living with or affected by HIV and AIDS?

Methods

Search Strategy and Procedure

Following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist, our search strategy protocol was published in PROSPERO (registration CRD42020161119). MEDLINE, PubMed, EMBASE, CINAHL, PsycINFO, CENTRAL (Cochrane Register of Controlled Trials), Scopus, and Google Scholar were searched electronically. The search strategy focused on 2 central concepts: the intervention (ie, online support groups) and the population (ie, individuals living with or affected by HIV and AIDS). It was developed using a combination of Medical Subject Headings and keywords with no study design filter. In addition, the references of the selected articles were hand searched for any additional relevant studies. The searches were conducted in April 2022 (for an example of the MEDLINE search and PRISMA checklist, see Multimedia Appendices 1 and 2).

Inclusion and Exclusion Criteria

To be eligible for inclusion in our scoping review, an article needed to (1) be peer-reviewed and (2) meet the inclusion criteria detailed in Textbox 1. These criteria were developed using the Population, Intervention, Comparator, Outcome, Setting, and Study Design model. We did not apply any restrictions with regard to date of publication. Our inclusion and exclusion criteria were developed such that we could focus solely on the unique contribution of text-based HIV and AIDS–related online support groups; therefore, we chose to exclude studies in which this support was combined with other forms of support.

Textbox 1. Population, Intervention, Comparator, Outcome, Setting, and Study Design inclusion and exclusion criteria.

Population

- Inclusion: individuals affected by HIV and AIDS either directly (ie, patient) or indirectly (eg, family member, loved one, colleague, or friend)
- Exclusion: individuals not affected by HIV and AIDS either directly or indirectly (eg, health professionals)

Intervention

- Inclusion: all types of support groups offered via the internet using either an asynchronous (eg, email listserv, message board, or open or closed social media groups) or synchronous (eg, chat room) text-based platform
- Exclusion: studies that evaluated a *combination* of face-to-face or telephone support with either an asynchronous or synchronous text-based platform

Comparator

- Inclusion: studies with or without a comparison or control group
- Exclusion: none

Outcome

- Inclusion: studies that reported on engagement with and utility of text-based HIV and AIDS–related online support groups in terms of psychosocial well-being and illness management, studies that reported the types of social support exchanged within text-based HIV and AIDS–related online support groups, and studies that reported negative experiences of engagement with text-based HIV and AIDS–related online support groups
- Exclusion: descriptive studies (eg, sociodemographic profile of users)

Setting

- Inclusion: text-based support platforms
- Exclusion: face-to-face setting

Study design

- Inclusion: all quantitative designs, qualitative studies that explored participants' experiences of online support groups as reported directly from the users, and studies reporting the analysis of user-generated content
- Exclusion: literature reviews and microanalysis of online discourse (eg, discourse or conversation analysis)

Study Selection

The 2 authors (NSC and HB) reviewed the titles and abstracts independently to identify potentially relevant articles. Abstracts not meeting the inclusion criteria were excluded. In cases where the abstract signaled potential eligibility, the full article was retrieved. Inclusion was based on agreement between both authors, and all reasons for exclusion were noted. In all instances of disagreement, discussion took place until the conflict was resolved.

Data Extraction and Data Synthesis

Predetermined study characteristics (ie, study aim, methods, data source, sample, and data analysis) as well as results (ie, quantitative findings, identified positive or negative outcomes, experiences or attributes, and types and frequencies of social support) were extracted by each of the 2 authors independently. To support this process, the review tool Covidence was used, and the extraction template was modified to support the specific requirements of the scoping review. Each of the 2 authors independently extracted the study characteristics and findings before discussing and agreeing on the final extraction content, which was then entered into a specific section of the Covidence data extraction template. To address each of the research questions, we used both narrative (ie, tabulation of findings from individual primary studies) and thematic (ie, an inductive approach to generate descriptive or analytical themes) synthesis.

Results

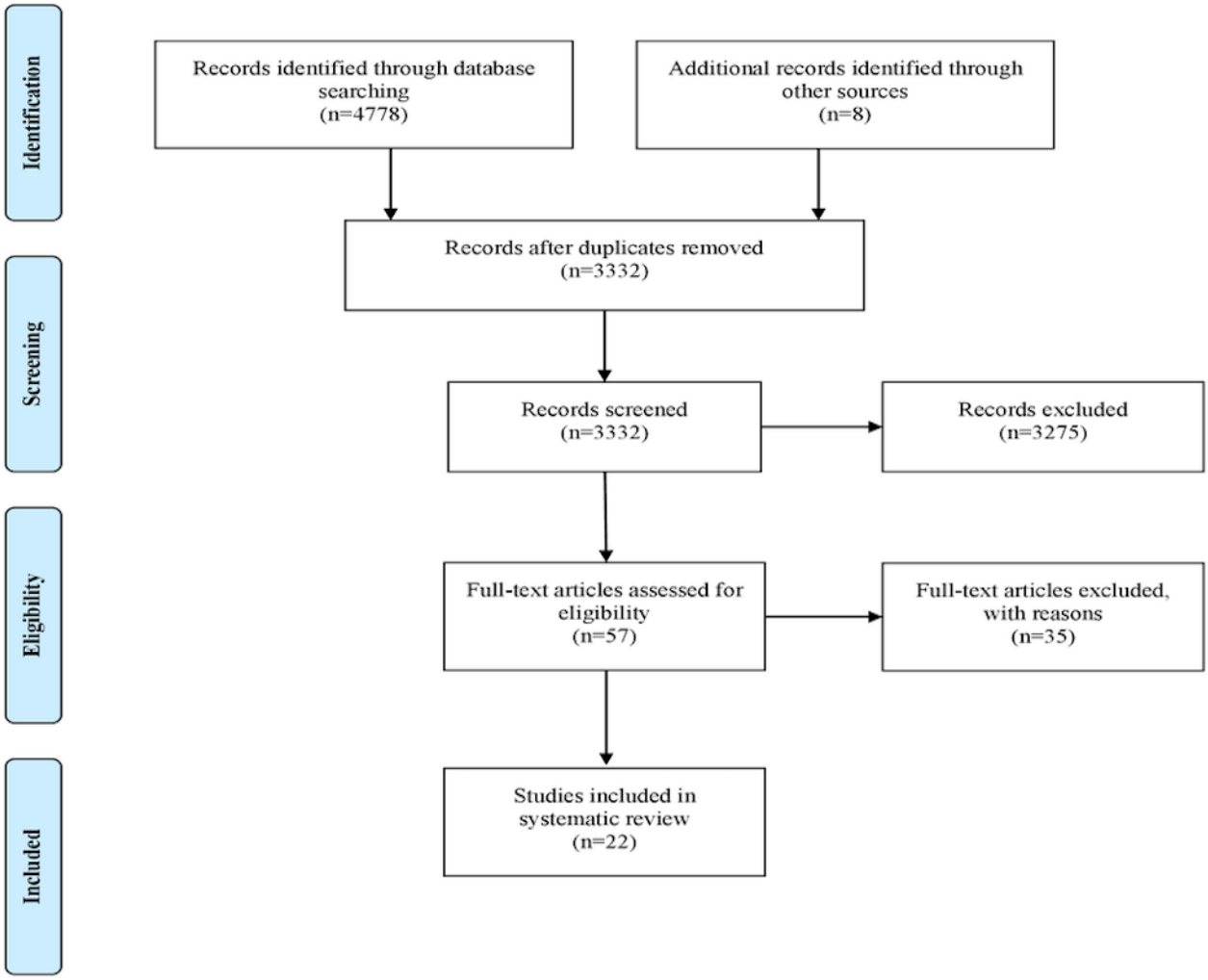
Included Studies

Our search strategy yielded 4786 studies, including 1454 (30.38%) duplicates. From the 3332 titles and abstracts reviewed, 2993 (89.83%) did not meet the inclusion criteria. The full texts of the remaining 57 studies were assessed, and a further 35 (61%) were excluded as they did not meet the inclusion criteria (Table 1) or were duplicates (n=11, 19%). Overall, there were 22 papers included in our review. Figure 1 presents the PRISMA flow diagram.

Table 1. Reasons for exclusion at the full-text review stage (N=35).

Reasons for exclusion	Studies, n (%)
No outcomes, experiences, or attributes of online support groups reported	14 (40)
Duplicates	11 (31)
Conference abstracts but that contained insufficient data for extraction	2 (6)
Microanalyses of online discourse	2 (6)
Peer support embedded within a complex intervention	2 (6)
Literature review or opinion article or discussion only	1 (3)
Not focused on HIV and AIDS	1 (3)
Focused on health professionals	1 (3)
Not obtainable	1 (3)

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of the studies included in the review.



Study Characteristics

For those studies included in the review, further details are provided in Table 2. As indicated, of the 22 studies, 5 (23%) were quantitative studies, 2 (9%) were mixed methods studies, and 15 (68%) were qualitative studies.

Table 2. Characteristics and details of the included studies (in alphabetical and date order).

Study	Aim (relevant to this review)	Intervention details	Design	Data collection	Sample	Analysis
Asiri et al [32]	To explore how social media users in the Arab world share sensitive health information through Facebook	Asynchronous forum, Facebook platform, Arabic only	Qualitative	Retrieval of user-generated messages	Not specified	Content analysis of 271 messages posted between December 1, 2010, and December 1, 2014
Bussone et al [33]	To investigate how people living with HIV and AIDS use a web-based community in terms of health information sharing	Public, asynchronous forum moderated by members	Qualitative	Retrieval of user-generated messages	Not specified	Thematic analysis of 252 messages
Coursaris and Liu [14]	To explore the exchange of social support within an HIV and AIDS–related online support group	Public, asynchronous forum	Qualitative	Retrieval of user-generated messages	Not specified	Content and thematic analysis of 5000 messages
Dong et al [34]	To explore whether social media could be used to study people living with HIV and AIDS and their needs and web-based habits	Asynchronous forum, Baidu Tieba platform, Chinese	Qualitative	Retrieval of user-generated messages	Not specified	Content analysis of 2340 messages posted between 2015 and 2017
Flickinger et al [35]	To examine how social support is exchanged as well as the benefits or limitations of a web-based community message board for people living with HIV	Private, asynchronous forum moderated by professionals	Qualitative	Retrieval of user-generated messages; semistructured interviews	55 HIV-positive individuals (mean age 39, SD 11.68 years; n=37, 67% male)	Content analysis of 840 messages and qualitative analysis of interviews
Flickinger et al [36]	To evaluate content posted to a web-based community message board for people living with HIV	Private, asynchronous forum moderated by professionals	Qualitative	Retrieval of user-generated messages	38 HIV-positive individuals (mean age 34.1, SD 11.5 years; n=28, 74% male)	Content analysis of 840 messages posted during the first 8 months of a community message board
Flickinger et al [37]	To understand the discussion of stigma within the web-based community message board and to evaluate participants' stigma levels at the 12-month follow-up	Private, asynchronous forum moderated by professionals	Mixed methods	Retrieval of user-generated messages; longitudinal survey	77 HIV-positive individuals (n=49, 64% male)	Content analysis of 394 messages; change in HIV-related stigma at 12 months
Gadgil et al [38]	To explore quality factors that may underpin the success of an online support group for people living with HIV	Private, asynchronous Facebook platform	Qualitative	Web-based semistructured interviews	32 HIV-positive individuals	Grounded theory analysis of interviews
Gaysynsky et al [39]	To examine the types of interaction that occurred within a Facebook online support group	Private, asynchronous forum, Facebook platform	Qualitative	Retrieval of user-generated messages	43 HIV-positive individuals (n=28, 65.1% male)	Content analysis of 3838 messages posted between March 1, 2011, and July 1, 2012
Guo and Goh [40]	To study changes in the composition of socioemotional and informational content in an HIV and AIDS online support group over time	Private, asynchronous, Sina Weibo microblogging platform moderated by the founding member	Qualitative	Retrieval of user-generated messages	Not specified	Content analysis of messages posted during the first 10 weeks (n=1277) of the online support group's existence and the last 10 weeks (n=966) of the data collection period

Study	Aim (relevant to this review)	Intervention details	Design	Data collection	Sample	Analysis
Han et al [41]	To explore how people who have self-labeled their HIV identity use social media	Public, asynchronous, Sina Weibo microblogging platform	Qualitative	Retrieval of user-generated messages	Not specified	Deductive thematic analysis of 1507 messages posted between January 1, 2015, and May 22, 2015
Han et al [42]	To examine whether individuals living with HIV and AIDS perceive higher levels of social support via Weibo than from offline relationships and whether perceived online support is associated with enacted social support and predictive of better health outcomes	Public, asynchronous, Sina Weibo microblogging platform	Quantitative	Cross-sectional web-based survey	432 HIV-positive individuals (mean age 29.2, SD 5.87 years; n=414, 95.8% male), years since HIV diagnosis (mean 2.69, range 0.1-11.4 years)	Multivariate analysis
Lai and Peirce [43]	To explore types of social support observed within public HIV and AIDS–related online support groups	Public, asynchronous forums (N=6)	Qualitative	Retrieval of user-generated messages	Not specified	Content analysis of 113 messages posted within the previous 30 days
Maestre et al [44]	To examine social support exchanges within HIV and AIDS–related online support groups	Public, asynchronous forums (N=4) moderated by professionals and members	Qualitative	Retrieval of user-generated messages	233 individuals	Content analysis of 400 messages (ie, the most recent 100 messages from each forum)
Mo and Coulson [45]	To examine the type of social support contained within messages posted to an HIV and AIDS–related online support group	Public, asynchronous forum (N=1)	Qualitative	Retrieval of user-generated messages	171 individuals	Content analysis of 85 threads randomly selected from all threads (n=342; 5230 messages) posted between June 1, 2006, and June 30, 2006
Mo and Coulson [46]	To explore whether differences exist between lurkers and posters in their use of HIV and AIDS–related online support groups, experience of empowering processes, and outcomes and satisfaction	Public, asynchronous forums (N=6), moderated	Quantitative	Cross-sectional web-based survey	340 HIV-positive individuals (mean age 47.8, SD 10.6 years; n=283, 83.7% male), years since HIV diagnosis (mean 11.8 years)	Multivariate analysis
Mo and Coulson [47]	To explore the use of online support groups and association with health status, coping, and social support among individuals living with HIV and AIDS	Asynchronous forum (N=1)	Quantitative	Cross-sectional web-based survey	640 HIV-positive individuals (mean age 45.52, SD 9.26 years; n=525, 82.4% male), years since HIV diagnosis (mean 9.69, SD 6.80 years)	Classification of online support group users into (1) nonusers, (2) infrequent users, and (3) frequent users followed by multivariate analysis
Mo and Coulson [48]	To explore the mechanisms through which participation in HIV and AIDS–related online support groups may promote patient empowerment	Public, asynchronous forums (N=6), moderated	Quantitative	Cross-sectional web-based survey	340 HIV-positive individuals (mean age 47.8, SD 10.6 years; n=283, 83.7% male), years since HIV diagnosis (mean 11.8 years)	Structural equation modeling

Study	Aim (relevant to this review)	Intervention details	Design	Data collection	Sample	Analysis
Mo and Coulson [49]	To examine the relationship among online support group use, patient empowerment, and psychological outcomes for individuals living with HIV and AIDS	Public, asynchronous forums (N=6), moderated	Quantitative	Cross-sectional web-based survey	340 HIV-positive individuals (mean age 47.8, SD 10.6 years; n=283, 83.7% male), years since HIV diagnosis (mean 11.8 years)	Structural equation modeling
Mo and Coulson [29]	To explore the presence of potentially empowering and disempowering processes and outcomes within HIV and AIDS–related online support groups	Public, asynchronous forums (N=4), moderated	Qualitative	Cross-sectional web-based survey	115 HIV-positive individuals (mean age 45.92, SD 9.96 years; n=102, 88.7% male), years since HIV diagnosis (mean 10.59, SD 13.77 years)	Thematic analysis of open-ended responses coding for both empowering and disempowering processes as well as outcomes
Peterson [50]	To explore how an HIV and AIDS–related online support group delivers positive social support and builds community	Private email listserv, moderated	Qualitative	Retrieval of user-generated messages	Not specified	Grounded theory analysis of 1870 messages posted over a 2-month period
Shi and Chen [51]	To explore the types of social support observed within a Chinese HIV and AIDS–related online support group	Public, asynchronous, Sina Weibo microblogging platform	Qualitative	Retrieval of user-generated messages	Not specified	Content analysis of all 7215 messages posted since its creation on January 18, 2011, as of September 14, 2012

Of the 6 quantitative designs (n=5, 83% quantitative studies and n=1, 17% mixed methods studies), only 1 (17%) was longitudinal and measured changes in HIV-related stigma over time [37]. The other 83% (5/6) of the quantitative studies were cross-sectional, measuring engagement with online support groups at only 1 time point [42,46-49]. The qualitative studies and the qualitative component of the mixed methods studies undertook an analysis of user-generated content [14,32-34,36,37,39-41,43-45,50,51] or interviewed online support group members [35,38], with 6% (1/17) of the studies analyzing the content of responses to open-ended questions [29].

As can be seen in Table 2, publication dates for the included studies ranged from 2007 to 2019; 45% (10/22) of the studies were published within the last 5 years, with 36% (8/22) being published within the last 10 years and 18% (4/22) being >10 years old. As judged by the address for the lead or corresponding author, the research teams were based in the United States (9/22, 41%), China or Hong Kong (5/22, 23%), the United Kingdom (4/22, 18%), Singapore (2/22, 9%), Australia (1/22, 5%), and Saudi Arabia (1/22, 5%). Online support groups included public and private asynchronous forums, microblogging websites, Facebook groups, and an email listserv. Across the 22 included studies, a total of 27,421 user-generated messages were analyzed along with 87 interviews and 1527 survey responses.

Synthesis of Results

Research Question 1: Does Engagement With Online Peer Support Groups Improve Psychosocial Well-being and Illness Management in Those Diagnosed With HIV and AIDS?

Overview

To address this research question, data were extracted and synthesized from 100% (5/5) of the quantitative studies [42,46-49] and the quantitative data from 50% (1/2) of the mixed methods studies [37]. These data are presented in Table 3 synthesized into a single theme (*Positive and negative associations between engagement and psychosocial outcomes*). Next, we extracted data from 27% (4/15) of the qualitative studies [14,35,38,39] and the qualitative data from 50% (1/2) of the mixed methods studies [29]. We synthesized this into a single theme (*Connecting with similar others*) capturing the positive impact of engagement on psychosocial well-being and illness management.

Table 3. Variables, measures, and key findings from the included quantitative and qualitative studies.

Study type and study	Variables and measures	Summary of key findings
Longitudinal		
Flickinger et al [37]	HIV-related stigma: 40-item Berger Stigma Scale (possible score range from 40 to 160)	Baseline: mean 102.94 (SD 18.26); 12 months: mean 98.73 (SD 15.08). There was a trend toward reduced stigma, with a mean change of −3.9 (95% CI −8.1 to 0.2), but it was not statistically significant (*P*=.06). Among those who posted, posters of content unrelated to stigma had a mean change in stigma scores of −3.3 (SD 12.7) compared with −5.1 (SD 17.2) for posters of stigma-related content. There was a trend toward more improvement in stigma scores with posting vs not posting and with posting about stigma vs other content, though these differences were not statistically significant (*P*=.50 and *P*=.72, respectively; 1-way ANOVA *F* test).
Cross-sectional		
Han et al [42]	OSG[a] use: length of use (years) and frequency of Weibo use; health information: date of HIV diagnosis, recent CD4 cell counts, and HIV status disclosure; enacted giving social support: 5 items; enacted receiving social support: 4 items; perceived offline social support: 12-item Multidimensional Scale of Perceived Social Support; perceived online social support: 8-item modified Multidimensional Scale of Perceived Social Support; subjective well-being: 5-item Satisfaction with Life Scale; adherence to ART[b]: 6 items; risky sexual behavior: 5 items	Perceived online social support was associated with employment status (employed people had higher levels of support compared with unemployed people), CD4 cell counts (those with lower CD4 cell counts perceived more support), and perceived offline support (higher offline support was associated with higher online support). People living with HIV and AIDS perceived higher levels of social support from Weibo than from offline support (from family and friends).
Mo and Coulson [46]	OSG use: length of use (months) and days and hours per average week; satisfaction with online support group experience: 4 items; empowering processes: 43-item scale, 4 processes (receiving useful information, receiving social support, finding positive meaning, and helping others); self-care self-efficacy: 29-item Strategies Used by People to Promote Health Scale; loneliness: 10-item UCLA[c] Loneliness Scale; optimism: 10-item Life Orientation Test-Revised; coping: 28-item Brief COPE[d]; depression: 20-item Center for Epidemiological Studies Depression Scale-Revised; quality of life: 35-item Medical Outcome Study HIV Health Survey	Compared with posters, members who only read the messages ("lurkers") scored lower in receiving social support and receiving useful information in empowering processes and lower in satisfaction with their relationship with group members. They also scored higher in distraction and lower in planning on the Brief COPE. In addition, they scored lower in social function and higher in energy. There were no significant differences in self-care self-efficacy, loneliness, depression, or optimism between posters and "lurkers."
Mo and Coulson [47]	Medical history: time since diagnosis, disease stage, and recent CD4 cell count; OSG use: hours in the previous month; health status: 36-item Medical Outcomes Study Short Form 36; coping: 28-item Brief COPE; perceived social support: 19-item Medical Outcomes Study Social Support Survey	Frequent users reported poorer health than nonusers. In addition, both frequent and infrequent users scored higher in planning, active coping, instrumental support, and emotional support coping on the Brief COPE. No significant difference was found for social support.
Mo and Coulson [48]	OSG use: length of use and days and hours per average week; empowering processes: 43-item scale, 4 processes (receiving useful information, receiving social support, finding positive meaning, and helping others); self-care self-efficacy: 29-item Strategies Used by People to Promote Health Scale; coping: 28-item Brief COPE; quality of life: 35-item Medical Outcome Study HIV Health Survey	The use of online support groups was significantly related to higher levels of all 4 empowering processes. Receiving useful information and finding positive meaning were related to higher levels of adaptive coping and lower levels of maladaptive coping, whereas receiving social support and helping others were related to higher levels of self-care self-efficacy, which in turn was related to higher levels of adaptive coping and lower levels of maladaptive coping. Finally, higher levels of adaptive coping and lower levels of maladaptive coping were related to better quality of life.
Mo and Coulson [49]	OSG use: length of use and days and hours per average week; empowering processes: 43-item scale, 4 processes (receiving useful information, receiving social support, finding positive meaning, and helping others); loneliness: 10-item UCLA Loneliness Scale; optimism: 10-item Life Orientation Test-Revised; depression: 20-item Center for Epidemiological Studies Depression Scale-Revised	Online support group use was positively related to empowering processes, which in turn was positively related to optimism toward life. Optimism was negatively related to loneliness and depression. Loneliness was also positively related to depression.
Qualitative		

Study type and study	Variables and measures	Summary of key findings
Coursaris and Liu [14]	N/A[e]	Members of the group shared their personal conditions, thoughts, and feelings related to HIV with others as well as expressing gratitude or sending congratulations. This disclosure and actions served to promote reciprocal disclosure and promote group ties. Consequently, members felt better about themselves.
Flickinger et al [35]	N/A	The community message board helped individuals connect with others going through a similar experience and fostered a sense of universality. The mutual exchange of support between members was also described as beneficial, with both informational and emotional support being particularly helpful in terms of outlook.
Gadgil et al [38]	N/A	Sharing experiences of stigma and memories of shame, guilt, and pain promoted a sense of camaraderie that mitigated the negative impacts of both felt and enacted stigma.
Gaysynsky et al [39]	N/A	Participation in an online support group helped members feel like they were being treated as an equal.
Mo and Coulson [29]	N/A	Engagement with online support groups was associated with six empowering processes: (1) exchanging information, (2) sharing experiences, (3) connecting with others, (4) encountering emotional support, (5) finding recognition and understanding, and (6) helping others. Six empowering outcomes arising from engagement with online support groups were identified: (1) increased optimism, (2) emotional well-being, (3) social well-being, (4) being better informed, (5) improved disease management, and (6) feeling confident in the relationship with physicians.

[a]OSG: online support group.

[b]ART: antiretroviral therapy.

[c]UCLA: University of California, Los Angeles.

[d]COPE: Coping Orientation to Problems Experienced.

[e]N/A: not applicable.

Positive and Negative Associations Between Engagement and Psychosocial Outcomes

Of the 6 quantitative studies, only 1 (17%) was a longitudinal investigation. Flickinger et al [37] examined changes in HIV-related stigma over time (ie, at baseline and after 12 months). Although there was a reduction in HIV-related stigma, this was not significant.

A total of 23% (5/22) of the studies adopted a cross-sectional approach when considering the relationship between engagement with online support groups and psychosocial well-being and illness management. Using the same data set, 9% (2/22) of the studies [46,49] found that engagement with online support groups was positively associated with greater exposure to a number of "empowering processes." These, in turn, were positively associated with adaptive coping, self-care self-efficacy, optimism, and quality of life [46] and lower levels of loneliness and depression [49]. A paper from this data set [46] considered whether these differences would exist when engagement was viewed in terms of "lurking" (ie, not posting any messages) versus "posting," but no differences were found in terms of self-care self-efficacy, loneliness, depression, or optimism. Similarly, no differences were found between frequent and infrequent users of online support groups in relation to perceived social support [47]. However, other sociodemographic and medical factors were predictive of perceived online support in 5% (1/22) of the studies [42]. Indeed, this study found that people living with HIV and AIDS perceived higher levels of social support from their online support group than from their offline networks.

Connecting With Similar Others

This theme captured the way in which engagement with HIV and AIDS–related online support groups was considered beneficial to individuals in terms of improving their psychosocial well-being and illness management. The qualitative studies illuminated how online support groups provided an opportunity to meet and interact with other individuals who were currently experiencing or had in the past experienced similar issues and challenges [14,29,35,38,39]. Through the sharing of personal experiences, thoughts, and feelings as well as messages offering congratulations, gratitude, and acknowledgment, a sense of camaraderie and group cohesion was developed. In addition, 5% (1/22) of the studies described how the exchange of informational and emotional support was helpful to group members [35]. Together, the ability to connect and engage in mutually supportive interaction helped individuals in terms of their self-worth [14,39] and outlook as well as mitigating the negative impacts of stigma [38]. A total of 5% (1/22) of the studies [29] described how this mutually supportive interaction also benefited individuals in terms of better illness

management and improved confidence when communicating with health care professionals.

Research Question 2: Are There Any Negative Aspects of Online Support Groups Experienced by Individuals Living With HIV and AIDS? If So, What Are These?

Overview

To address this research question, data were extracted and synthesized from 33% (5/15) of the qualitative studies [29,32,35,36,39] and the qualitative data from 50% (1/2) of the mixed methods studies [37]. Three themes capturing the negative aspects of online support groups experienced by users were generated from the findings (Table 4): (1) challenging behavior, (2) difficult content, and (3) negative consequences of the web-based platform.

Table 4. Qualitative and descriptive findings regarding the negative aspects of engagement.

Study	Negative aspects of engagement: themes
Asiri et al [32]	• Negative judgment or attribution of blame • Concerns about privacy and disclosure
Flickinger et al [35]	• Challenging or negative content (complaining, suicidal ideation, attacking, vulgarity, poor language, taboo topics, excessive personal information, and religiosity) • Feelings of obligation or keeping up to date with conversations • Lack of activity or immediacy of feedback • Lack of anonymity • Being unable to connect physically • Feeling like an outsider in the community • Potential of lost relationships once the research study was completed
Flickinger et al [36]	• The impact of negative posts—disturbing or disruptive to the community (eg, posts expressing suicidal thoughts or mental health concerns)
Flickinger et al [37]	• Negative thread interactions—contained posts expressing strong emotions on the CMBa or posts containing a negative reaction to another member's post
Gaysynsky et al [39]	• Interpersonal conflict (disrespectful, sarcastic, unkind, or argumentative); statements that expressed being hurt, distressed, or angered by other members of the group and statements that demonstrated disagreement, tension, or antagonism
Mo and Coulson [29]	• Being unable to connect physically • Inappropriate behavior on the web (inappropriate, disrespectful, attacking, or ridiculing) • Declining real-life relationships (overreliance on web-based relationships) • Informational overload and misinformation

[a]CMB: community message board.

Challenging Behavior

This theme was concerned with behavior of group members that was considered inappropriate. Some studies (2/6, 33%) described how members felt that the behavior of others was inappropriate, unkind, disrespectful, or attacking or was trying to ridicule others [29,39].

Negative Content

Several studies (4/6, 67%) reported how the content posted to the online support groups was negative or difficult to deal with. For example, Flickinger et al [35] reported instances of complaining, vulgarity, bad language, taboo topics, excessive personal information, and religiosity. Furthermore, the challenge of being exposed to posts that were disturbing or disruptive to the community, such as posts expressing suicidal thoughts or mental health concerns, was highlighted [36]. Asiri et al [32] also reported instances in which some of the content posted on the web appeared to be judging or ascribing blame to individuals for their HIV status.

Negative Consequences of the Web-Based Platform

A range of issues was identified that reflected the negative consequences arising from the platform underpinning the online support group. For example, 33% (2/6) of the studies noted the difficulties arising from the fact that group members were not physically copresent [35], and this presented challenges in terms of forming web-based relationships [29]. However, some concerns were also expressed regarding overreliance on these web-based relationships and the potential for a decline in real-world relationships as a possible consequence [29]. Concerns were also expressed regarding the quantity and quality of information exchanged between group members. For example, Flickinger et al [35] noted the difficulties experienced by individuals when trying to keep up with conversations on the web, whereas other studies (1/6, 17%) reported difficulties regarding both information overload and accuracy [29]. Flickinger et al [35] described problems concerning a lack of activity or immediacy of feedback on members' posts. In addition, they noted concerns regarding the lack of anonymity

in group participation. Asiri et al [32] also identified concerns regarding privacy and disclosure in their study.

Research Question 3: What Types of Social Support Are Exchanged Within Online Support Groups for Individuals Affected by HIV or AIDS?

To address this research question, data were extracted and synthesized from 73% (11/15) of the qualitative studies [14,33-35,39-41,43-45,51], which analyzed user-generated content for evidence of social support exchange. These are presented in Table 5. Most studies (8/11, 73%) included sufficient detail to populate Table 5 but, in some instances, we undertook a simple calculation from the available data to report support requests or provision.

Table 5. Social support exchange.

Study	Overall	Social support requests (% of total requests)	Social support provision (% of total provision)
Bussone et al [33]	• 252/2455 (10.26%) messages either asked or responded to questions about personal health information; 60/2455 (2.4%) messages requested social support; 192/2455 (7.8%) messages provided social support	• 60/2455 (2.4%) messages that contained 77 different questions	• Informational support: 176/192 (91.7%) • Emotional support: 56/192 (29.2%) • Esteem support: 15/192 (7.8%) • Network support: 15/192 (7.8%) • Tangible support: 1/192 (0.5%)
Coursaris and Liu [14]	• 815/5000 (16.3%) messages requested social support; 2310/5000 (41.6%) messages provided social support • 87% of messages included 1 type of support, 12% included 2 types, and 1% included 3 types	• Informational support: 626/815 (76.8%) • Emotional support: 154/815 (18.9%) • Network support: 78/815 (9.6%) • Esteem support: 26/815 (3.2%) • Tangible assistance: 13/815 (1.6%)	• Informational support: 1458/2310 (63.1%) • Emotional support: 646/2310 (28%) • Esteem support: 294/2310 (12.7%) • Network support: 260/2310 (11.3%) • Tangible assistance: 27/2310 (1.2%)
Dong et al [34]	• 726/2340 messages (31.03%) included social support; 559/2340 (23.9%) messages requested social support; 167/2340 (7.1%) messages provided social support	• Request for friendship: 436/559 (78%)	• Sharing knowledge: 43/167 (25.8%)
Flickinger et al [35]	• 115/840 (14%) messages requested social support; 433/840 (52%) messages provided social support	• Emotional support: 85/840 (10.1%) • Informational support: 30/840 (3.6%) • No evidence of instrumental support	• Emotional support: 178/433 (41.1%) • Network support: 115/433 (26.6%) • Esteem support: 77/433 (17.8%) • Informational support: 55/433 (12.7%) • Instrumental support: 8/433 (1.8%)
Gaysynsky et al [39]	• 255/3838 (6.6%) messages requested social support; 578/3838 (15.1%) messages provided social support	• Emotional support: 82/255 (32.2%) • Network support: 80/255 (31.4%) • Informational support: 56/255 (22%) • Tangible assistance: 37/255 (14.5%) • Esteem support: 21/255 (8.2%)	• Esteem support: 259/578 (44.8%) • Emotional support: 149/578 (25.8%) • Network support: 113/578 (19.6%) • Informational support: 83/578 (14.4%)
Guo and Goh [40]	• 1277 messages posted during the first 10 weeks; 966 messages posted during the last 10 weeks • Socioemotional messages—481 (37.67%) to 494 (51.14%)—exceeded informational messages—796 (62.33%) to 472 (48.86%)—over time	—[a]	—
Han et al [41]	• 135/1507 (9%) messages requested social support; 603/1507 (40%) messages provided social support	• Emotional support: 57/135 (42.2%) • Informational support: 52/135 (38.5%) • Others: 26/135 (19.3%)	• HIV and AIDS–related: 104/603 (17.2%) • Daily life events: 499/603 (82.7%)

Study	Overall	Social support requests (% of total requests)	Social support provision (% of total provision)
Lai and Peirce [43]	• 11/113 (9.7%) messages requested social support; 104/113 (92%) messages provided social support	• Emotional support: 5/11 (45.5%) • Informational support: 3/11 (2.7%) • Instrumental support: 3/11 (2.7%)	• Informational support: 61/104 (58.7%) • Emotional support: 34/104 (32.7%) • Instrumental support: 9/104 (8.7%)
Maestre et al [44]	• 400 messages yielding 525 utterances; 292/525 (55.6%) utterances requested social support; 233/525 (44.4%) utterances provided social support	• Informational support: 165/292 (56.5%) • Emotional support: 108/292 (37%) • Esteem support: 13/292 (4.5%) • Network support: 6/292 (2.1%) • No evidence of tangible support	• Informational support: 126/233 (54.1%) • Emotional support: 83/233 (35.6%) • Esteem support: 17/233 (7.3%) • Network support: 7/233 (3%) • No evidence of tangible support
Mo and Coulson [45]	• 986/1138 (86.6%) messages contained at least one type of social support	—	• Informational support: 466/986 (47.3%) • Emotional support: 369/986 (37.4%) • Esteem support: 130/986 (13.2%) • Network support: 72/986 (7.3%) • Tangible assistance: 10/986 (1%)
Shi and Chen [51]	• 5589/7215 (77.46%) messages included social support; 1588/7215 (22.01%) messages requested social support; 4001/7215 (55.45%) messages provided social support	• Informational support: 1051/1588 (66.2%) • Instrumental support: 319/1588 (20.1%) • Emotional support: 218/1588 (13.7%)	• Informational support: 2528/4001 (63.2%) • Emotional support: 1405/4001 (35.1%) • Instrumental support: 68/4001 (1.7%)

[a]Not available.

All studies used the message as the level of analysis except for the study by Maestre et al [44], which used the utterance. All studies (11/11, 100%) used a deductive analytical approach, with 55% (6/11) of the studies [14,33,35,39,44,45] using the Social Support Behavior Code developed by Cutrona and Suhr [52] as their underpinning coding framework. From the remaining studies, 9% (1/11) [41] combined the Social Support Behavior Code with interaction process analysis, 9% (1/11) [43] used a social support conceptual framework developed by House and Kahn [53], and 9% (1/11) [51] were guided by the typology described by Wright et al [54]. A further 18% (2/11) of the studies [34,41] stated that researchers with experience in public health developed the categories of social support to be coded.

A total of 82% (9/11) of the studies distinguished between social support requests and provision [14,33-35,39,41,43,44,51]. Of these 9 studies, 2 (22%) [34,44] reported a higher percentage of support requests than support provision, whereas 7 (78%) [14,33,35,39,41,43,51] reported a higher percentage of messages classified as providing social support. In terms of support requests, 36% (4/11) of the studies reported emotional support as being the most frequently requested [35,39,41,43], and 27% (3/11) reported informational support as being the most frequently requested [14,44,51]. For support provision, 55% (6/11) of the studies [14,33,43-45,51] reported informational support as being the most prevalent type of social support offered, with 9% (1/11) [35] reporting emotional support as being the most common and 9% (1/11) [39] reporting esteem support as being the most common. Only Guo and Goh [40] considered how the exchange of social support changed over time. They found that socioemotional messages exceeded informational support messages over time.

Discussion

Principal Findings

Online support groups offer new opportunities for those living with HIV and AIDS to access information, advice, and mutual peer support. To our knowledge, this is the first scoping review to synthesize the evidence regarding HIV or AIDS–related online support groups. Our review had three aims; we sought to determine (1) whether engagement with online peer support groups improved psychosocial well-being and illness management for those living with HIV and AIDS; (2) whether there existed any negative aspects of online support groups experienced by individuals living with or affected by HIV and AIDS and, if so, what were they; and (3) what types of social support were exchanged within online support groups for individuals living with or affected by HIV and AIDS. We addressed the first research question by synthesizing the findings

of a range of published quantitative, mixed methods, and qualitative studies. The second and third research questions were answered by synthesizing the findings of published qualitative studies.

In terms of our first research question, no randomized controlled trials of the impact of engagement with HIV and AIDS–related online support groups on psychosocial well-being and illness management were identified. However, we did extract relevant findings from quantitative and qualitative studies as well as the quantitative component of a mixed methods study. There was limited quantitative evidence that engagement with online peer support groups improves psychosocial well-being. Indeed, the only longitudinal study conducted over a 12-month period [37] reported no changes in HIV-related stigma. All the cross-sectional studies (5/22, 23%) reported an association between engagement with HIV or AIDS–related online support groups and psychosocial well-being and illness management [42,46-49]; however, it is difficult to make any causal inference owing to the cross-sectional nature of the studies. In terms of qualitative evidence, benefits arising from engagement with HIV and AIDS–related online support groups were reported across all the studies (15/22, 68%), in particular the psychosocial benefits of individuals coming together on the web and sharing their experiences together with the mutual exchange of support. Mo and Coulson [29] described a range of "active ingredients" that may benefit individuals who engage with HIV or AIDS–related online support groups. These included exchanging information, sharing experiences, connecting with others, encountering emotional support, finding recognition, and understanding and helping others. In turn, they identified several psychosocial outcomes that may arise from engagement, including increased optimism and control over the future, improved emotional and social well-being, being better informed, improved coping, and feeling more confident in their relationship with health care professionals.

In relation to our second research question, there were 27% (6/22) of studies reporting qualitative data on the negative aspects of engagement with online support groups. We identified 3 key problematic issues. The first revolved around the challenging behavior of other group members, particularly in terms of interpersonal conflict. Next, we identified the challenge of negative content and how group members could be exposed to online material that was difficult to read. Finally, we noted negative experiences that may arise from the unique features (eg, text-based and asynchronous) of the web-based platforms used to support the online support groups. Specifically, we noted difficulties in forming online relationships but also instances of overreliance on these relationships, which may then negatively affect offline relationships. We also identified concerns regarding both information quantity and quality. These various concerns have been reported elsewhere in the literature [24,27,55] and are potentially serious in nature. However, we currently have little understanding of the long-term impact of these negative experiences on group members. Moving forward, these findings do suggest that online support group moderators or administrators may play a crucial role in achieving the aims of the support group and safeguarding its membership.

Half of the studies included in our review (11/22, 50%) addressed the third research question, which considered the types of social support exchanged within online support groups for individuals living with or affected by HIV and AIDS. In total, 100% (11/11) of the qualitative studies used a deductive analytic approach with a social support framework (or typology) to guide the analysis, with most studies (6/11, 55%) using the Social Support Behavior Code developed by Cutrona and Suhr [52]. There were more studies reporting a higher proportion of support provision than of requests. This may be explained by the asynchronous text-based platforms that were used in the studies. On these platforms, conversational threads are developed by individual group members posting a message and other group members posting replies. Our findings also revealed that emotional and informational support were the types most frequently requested but, in terms of provision, informational support was the most common type of support reported. These findings are consistent with the results of a meta-analytic review of 41 published studies that reported informational and emotional support messages as the most prevalent within health-related online support groups [56].

Strengths and Limitations of This Review

It is important to highlight the strengths of this scoping review. Most notably, we identified, described, and synthesized data from quantitative, qualitative, and mixed methods studies and considered the guidance set out by Booth et al [57] to support this process. In doing so, we believe this has provided the reader with a richer and more holistic insight into the role of online support groups for individuals affected by HIV and AIDS. However, there are also some limitations that should be considered. First, although we searched 7 databases, it is possible that we failed to identify some relevant studies. However, to mitigate this, we also searched Google Scholar as well as hand searching the reference lists of all the included studies. This yielded additional studies that were included in our review. Second, our review may have introduced bias through the inclusion of only studies published in English; thus, we may have overlooked studies published in other languages. Finally, our review identified a few quantitative studies, and those that were included were mostly cross-sectional surveys. Therefore, it becomes difficult to draw definitive conclusions, particularly in relation to the psychosocial benefits arising from engagement with HIV and AIDS–related online support groups.

Recommendations for Future Research

This scoping review has revealed that there exists a growing and diverse body of literature that considers the role of online support groups for people affected by HIV and AIDS. This literature includes quantitative, qualitative, and mixed methods research designs. However, our review also pinpointed specific areas for future research to advance our understanding of the role and impact of online support groups for individuals affected by HIV and AIDS. First, to assess the impact of engagement more fully, future research efforts should seek to develop more robust research designs, including randomized controlled trials and longitudinal studies on both the benefits and harms of engagement. This work should also seek to consider a broader array of psychosocial as well as illness-related outcome

measures. Second, as these online support groups may be supported by a range of platforms, future research should seek to explore how the affordances of each platform may influence both engagement and outcomes. Third, our review noted that online support group moderators may play a pivotal role in promoting the aims of the group as well as safeguarding its membership. Therefore, further research examining the function and effectiveness of moderator web-based behavior is warranted. Fourth, as the exchange of informational support appears widespread within online support groups for HIV and AIDS, future work may seek to determine the accuracy of any medical-related information shared and the extent to which it may affect the coping strategies and behaviors of members.

Conclusions

Online support groups provide an opportunity for individuals affected by HIV and AIDS to engage in mutual support. This engagement may be associated with improved illness management as well as a range of beneficial psychosocial outcomes. However, members may experience negative aspects of the online support groups, particularly in terms of interpersonal conflict with other members and content that is challenging. Online support groups for HIV and AIDS can provide a valuable opportunity to both seek and provide social support, notably informational and emotional support.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Example of MEDLINE search strategy from April 2022.
[DOC File , 43 KB - jmir_v24i7e27648_app1.doc]

Multimedia Appendix 2
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 108 KB - jmir_v24i7e27648_app2.docx]

References

1. HIV/AIDS Fact Sheet. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/hiv-aids [accessed 2022-04-01]
2. Hudson A, Lee K, Miramontes H, Portillo C. Social interactions, perceived support, and level of distress in HIV-positive women. J Assoc Nurses AIDS Care 2001 Jul;12(4):68-76 [FREE Full text] [doi: 10.1016/s1055-3290(06)60218-5]
3. Colbert A, Kim K, Sereika S, Erlen J. An examination of the relationships among gender, health status, social support, and HIV-related stigma. J Assoc Nurses AIDS Care 2010;21(4):302-313 [FREE Full text] [doi: 10.1016/j.jana.2009.11.004] [Medline: 20116295]
4. Varni S, Miller C, McCuin T, Solomon S. Disengagement and engagement coping with HIV/AIDS stigma and psychological well-being of people with HIV/AIDS. J Soc Clin Psychol 2012 Feb 01;31(2):123-150 [FREE Full text] [doi: 10.1521/jscp.2012.31.2.123] [Medline: 22611302]
5. Turan B, Budhwani H, Fazeli PL, Browning W, Raper J, Mugavero M, et al. How does stigma affect people living with HIV? The mediating roles of internalized and anticipated HIV stigma in the effects of perceived community stigma on health and psychosocial outcomes. AIDS Behav 2017 Jan;21(1):283-291 [FREE Full text] [doi: 10.1007/s10461-016-1451-5] [Medline: 27272742]
6. Greene M, Hessol N, Perissinotto C, Zepf R, Hutton Parrott A, Foreman C, et al. Loneliness in older adults living with HIV. AIDS Behav 2018 May;22(5):1475-1484 [FREE Full text] [doi: 10.1007/s10461-017-1985-1] [Medline: 29151199]
7. Flowers P, Davis M, Larkin M, Church S, Marriott C. Understanding the impact of HIV diagnosis amongst gay men in Scotland: an interpretative phenomenological analysis. Psychol Health 2011 Oct;26(10):1378-1391 [FREE Full text] [doi: 10.1080/08870446.2010.551213] [Medline: 22010635]
8. Nachega J, Morroni C, Zuniga J, Sherer R, Beyrer C, Solomon S, et al. HIV-related stigma, isolation, discrimination, and serostatus disclosure: a global survey of 2035 HIV-infected adults. J Int Assoc Physicians AIDS Care (Chic) 2012;11(3):172-178 [FREE Full text] [doi: 10.1177/1545109712436723] [Medline: 22431893]
9. Spaan P, van Luenen S, Garnefski N, Kraaij V. Psychosocial interventions enhance HIV medication adherence: a systematic review and meta-analysis. J Health Psychol 2020 Sep;25(10-11):1326-1340 [FREE Full text] [doi: 10.1177/1359105318755545] [Medline: 29417851]
10. Vanable P, Carey M, Blair D, Littlewood R. Impact of HIV-related stigma on health behaviors and psychological adjustment among HIV-positive men and women. AIDS Behav 2006 Sep;10(5):473-482 [FREE Full text] [doi: 10.1007/s10461-006-9099-1] [Medline: 16604295]
11. Cooperman N, Simoni J. Suicidal ideation and attempted suicide among women living with HIV/AIDS. J Behav Med 2005 Apr;28(2):149-156 [FREE Full text] [doi: 10.1007/s10865-005-3664-3] [Medline: 15957570]

12. Bhatia M. Prevalence of depression in people living with HIV/AIDS undergoing ART and factors associated with it. J Clin Diagnostic Res 2014:WC01-WC04 [FREE Full text] [doi: 10.7860/jcdr/2014/7725.4927]
13. Anagnostopoulos A, Ledergerber B, Jaccard R, Shaw SA, Stoeckle M, Bernasconi E, Swiss HIV Cohort Study. Frequency of and risk factors for depression among participants in the Swiss HIV Cohort Study (SHCS). PLoS One 2015 Oct 22;10(10):e0140943 [FREE Full text] [doi: 10.1371/journal.pone.0140943] [Medline: 26492488]
14. Coursaris C, Liu M. An analysis of social support exchanges in online HIV/AIDS self-help groups. Comput Human Behav 2009 Jul;25(4):911-918 [FREE Full text] [doi: 10.1016/j.chb.2009.03.006]
15. Turner-Cobb J, Gore-Felton C, Marouf F, Koopman C, Kim P, Israelski D, et al. Coping, social support, and attachment style as psychosocial correlates of adjustment in men and women with HIV/AIDS. J Behav Med 2002 Aug;25(4):337-353. [doi: 10.1023/a:1015814431481] [Medline: 12136496]
16. Ashton E, Vosvick M, Chesney M, Gore-Felton C, Koopman C, O'Shea K, et al. Social support and maladaptive coping as predictors of the change in physical health symptoms among persons living with HIV/AIDS. AIDS Patient Care STDS 2005 Sep;19(9):587-598 [FREE Full text] [doi: 10.1089/apc.2005.19.587] [Medline: 16164385]
17. Internet World Stats homepage. Internet World Stats. 2022 Apr 01. URL: https://www.internetworldstats.com/ [accessed 2022-07-06]
18. What is HINTS? National Cancer Institute Health Information National Trends Survey. URL: http://hints.cancer.gov [accessed 2022-04-01]
19. O'Neill B, Ziebland S, Valderas J, Lupiáñez-Villanueva F. User-generated online health content: a survey of internet users in the United Kingdom. J Med Internet Res 2014 Apr 30;16(4):e118 [FREE Full text] [doi: 10.2196/jmir.3187] [Medline: 24784798]
20. Wright K. Communication in health-related online social support groups/communities: A review of research on predictors of participation, applications of social support theory, and health outcomes. RCR 2016;4:65-87 [FREE Full text] [doi: 10.12840/issn.2255-4165.2016.04.01.010]
21. Vilhauer R, McClintock M, Matthews A. Online support groups for women with metastatic breast cancer: a feasibility pilot study. J Psychosoc Oncol 2010;28(5):560-586 [FREE Full text] [doi: 10.1080/07347332.2010.504504] [Medline: 20730665]
22. Idriss SZ, Kvedar JC, Watson AJ. The role of online support communities: benefits of expanded social networks to patients with psoriasis. Arch Dermatol 2009 Jan 01;145(1):46-51. [doi: 10.1001/archdermatol.2008.529] [Medline: 19153342]
23. Coulson N, Knibb R. Coping with food allergy: exploring the role of the online support group. Cyberpsychol Behav 2007 Feb;10(1):145-148 [FREE Full text] [doi: 10.1089/cpb.2006.9978] [Medline: 17305463]
24. Coulson N. How do online patient support communities affect the experience of inflammatory bowel disease? An online survey. JRSM Short Rep 2013;4(8):2042533313478004 [FREE Full text] [doi: 10.1177/2042533313478004] [Medline: 24040493]
25. Buchanan H, Coulson N. Accessing dental anxiety online support groups: an exploratory qualitative study of motives and experiences. Patient Educ Couns 2007 Jun;66(3):263-269 [FREE Full text] [doi: 10.1016/j.pec.2006.12.011] [Medline: 17320336]
26. Davison K, Pennebaker J, Dickerson S. Who talks? The social psychology of illness support groups. Am Psychol 2000;55(2):205-217 [FREE Full text] [doi: 10.1037/0003-066x.55.2.205]
27. Coulson NS, Bullock E, Rodham K. Exploring the therapeutic affordances of self-harm online support communities: an online survey of members. JMIR Ment Health 2017 Oct 13;4(4):e44 [FREE Full text] [doi: 10.2196/mental.8084] [Medline: 29030324]
28. Pfeil U, Zaphiris P, Wilson S. Older adults' perceptions and experiences of online social support. Interact Comput 2009 Jul;21(3):159-172 [FREE Full text] [doi: 10.1016/j.intcom.2008.12.001]
29. Mo P, Coulson N. Are online support groups always beneficial? A qualitative exploration of the empowering and disempowering processes of participation within HIV/AIDS-related online support groups. Int J Nurs Stud 2014 Jul;51(7):983-993 [FREE Full text] [doi: 10.1016/j.ijnurstu.2013.11.006] [Medline: 24342449]
30. Broom A. The eMale. J Sociol 2016 Jun 30;41(1):87-104 [FREE Full text] [doi: 10.1177/1440783305050965]
31. Harding C, Chung H. Behavioral health support and online peer communities: international experiences. Mhealth 2016 Nov 23;2:43 [FREE Full text] [doi: 10.21037/mhealth.2016.10.04] [Medline: 28293613]
32. Asiri E, Khalifa M, Shabir S, Hossain MN, Iqbal U, Househ M. Sharing sensitive health information through social media in the Arab world. Int J Qual Health Care 2017 Feb 01;29(1):68-74. [doi: 10.1093/intqhc/mzw137] [Medline: 28003369]
33. Bussone A, Stumpf S, Wilson S. The use of online forums by people living with HIV for help in understanding personal health information. Int J Med Inform 2017 Dec;108:64-70 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.10.001] [Medline: 29132633]
34. Dong Y, Zhou X, Lin Y, Pan Q, Wang Y. HIV-related posts from a Chinese internet discussion forum: an exploratory study. PLoS One 2019;14(2):e0213066 [FREE Full text] [doi: 10.1371/journal.pone.0213066] [Medline: 30818379]
35. Flickinger T, DeBolt C, Waldman A, Reynolds G, Cohn W, Beach M, et al. Social support in a virtual community: analysis of a clinic-affiliated online support group for persons living with HIV/AIDS. AIDS Behav 2017 Nov;21(11):3087-3099 [FREE Full text] [doi: 10.1007/s10461-016-1587-3] [Medline: 27766448]

36. Flickinger T, DeBolt C, Wispelwey E, Laurence C, Plews-Ogan E, Waldman A, et al. Content analysis and user characteristics of a smartphone-based online support group for people living with HIV. Telemed J E Health 2016 Sep;22(9):746-754 [FREE Full text] [doi: 10.1089/tmj.2015.0160] [Medline: 27002956]
37. Flickinger T, DeBolt C, Xie A, Kosmacki A, Grabowski M, Waldman A, et al. Addressing stigma through a virtual community for people living with HIV: a mixed methods study of the positivelinks mobile health intervention. AIDS Behav 2018 Oct;22(10):3395-3406 [FREE Full text] [doi: 10.1007/s10461-018-2174-6] [Medline: 29882048]
38. Gadgil G, Prybutok G, Prybutok V. Qualitative investigation of the role of quality in online community support for people living with HIV and AIDS. Quality Manag J 2018 Oct 18;25(4):171-185 [FREE Full text] [doi: 10.1080/10686967.2018.1515524]
39. Gaysynsky A, Romansky-Poulin K, Arpadi S. "My YAP Family": analysis of a Facebook group for young adults living with HIV. AIDS Behav 2015 Jun;19(6):947-962 [FREE Full text] [doi: 10.1007/s10461-014-0887-8] [Medline: 25186783]
40. Guo Y, Goh DH. Corrigendum to ""I Have AIDS": content analysis of postings in HIV/AIDS support group on a Chinese microblog" [Comput. Hum. Behav. 34 (2014) 219–226]. Comput Human Behav 2015 Dec;53:219-226 [FREE Full text] [doi: 10.1016/j.chb.2015.06.008]
41. Han J, Tian X, Yu G, He F. Disclosure pattern of self-labeled people living with HIV/AIDS on Chinese social networking site: an exploratory study. Cyberpsychol Behav Soc Netw 2016 Aug;19(8):516-523. [doi: 10.1089/cyber.2016.0133] [Medline: 27494331]
42. Han X, Li B, Qu J, Zhu Q. Weibo friends with benefits for people live with HIV/AIDS? The implications of Weibo use for enacted social support, perceived social support and health outcomes. Soc Sci Med 2018 Aug;211:157-163 [FREE Full text] [doi: 10.1016/j.socscimed.2018.06.016] [Medline: 29940485]
43. Lai T, Peirce A. Analyzing online public support groups for people living with HIV/AIDS. AMIA Annu Symp Proc 2007 Oct 11:1022. [Medline: 18694120]
44. Maestre J, Herring S, Min A, Connelly C, Shih P. Where and how to look for help matters: analysis of support exchange in online health communities for people living with HIV. Inform 2018 Oct 20;9(10):259 [FREE Full text] [doi: 10.3390/info9100259]
45. Mo P, Coulson N. Exploring the communication of social support within virtual communities: a content analysis of messages posted to an online HIV/AIDS support group. Cyberpsychol Behav 2008 Jun;11(3):371-374 [FREE Full text] [doi: 10.1089/cpb.2007.0118] [Medline: 18537512]
46. Mo P, Coulson N. Empowering processes in online support groups among people living with HIV/AIDS: a comparative analysis of 'lurkers' and 'posters'. Comput Human Behav 2010 Sep;26(5):1183-1193 [FREE Full text] [doi: 10.1016/j.chb.2010.03.028]
47. Mo P, Coulson N. Living with HIV/AIDS and use of online support groups. J Health Psychol 2010 Apr;15(3):339-350 [FREE Full text] [doi: 10.1177/1359105309348808] [Medline: 20348355]
48. Mo P, Coulson N. Developing a model for online support group use, empowering processes and psychosocial outcomes for individuals living with HIV/AIDS. Psychol Health 2012;27(4):445-459 [FREE Full text] [doi: 10.1080/08870446.2011.592981] [Medline: 21854088]
49. Mo P, Coulson N. Online support group use and psychological health for individuals living with HIV/AIDS. Patient Educ Couns 2013 Dec;93(3):426-432 [FREE Full text] [doi: 10.1016/j.pec.2013.04.004] [Medline: 23669152]
50. Peterson J. "You have to be positive." social support processes of an online support group for men living with HIV. Commun Stud 2009 Oct 21;60(5):526-541 [FREE Full text] [doi: 10.1080/10510970903260368]
51. Shi J, Chen L. Social support on Weibo for people living with HIV/AIDS in China: a quantitative content analysis. Chin J Commun 2014 Jul 24;7(3):285-298 [FREE Full text] [doi: 10.1080/17544750.2014.926954]
52. Cutrona CE, Suhr JA. Controllability of stressful events and satisfaction with spouse support behaviors. Commun Res 2016 Jun 30;19(2):154-174 [FREE Full text] [doi: 10.1177/009365092019002002]
53. House, J., Kahn, R.. In: Cohen S, Syme L, editors. Social Support and Health. San Francisco: Academic Press; 1985:83-103.
54. Wright, K., Johnson, A., Bernard, D., Averbeck, J.. Computer-mediated social support. In: Thompson T, Parrott R, Nussbaum J, editors. The Routledge Handbook of Health Communication. New York: Routledge; 2011:349-362.
55. Shoebotham A, Coulson N. Therapeutic affordances of online support group use in women with endometriosis. J Med Internet Res 2016 May 09;18(5):e109 [FREE Full text] [doi: 10.2196/jmir.5548] [Medline: 27160641]
56. Rains S, Peterson E, Wright K. Communicating social support in computer-mediated contexts: a meta-analytic review of content analyses examining support messages shared online among individuals coping with illness. Commun Monographs 2015 Mar 17;82(4):403-430 [FREE Full text] [doi: 10.1080/03637751.2015.1019530]
57. Guidance on choosing qualitative evidence synthesis methods for use in health technology assessments of complex interventions. INTEGRATE-HTA. 2016. URL: http://www.integrate-hta.eu/downloads/) [accessed 2022-04-01]

Abbreviations

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

Edited by A Mavragani; submitted 01.02.21; peer-reviewed by G Rouleau, R Poluru; comments to author 23.05.21; revised version received 29.09.21; accepted 29.04.22; published 26.07.22.

Please cite as:
Coulson NS, Buchanan H
The Role of Online Support Groups in Helping Individuals Affected by HIV and AIDS: Scoping Review of the Literature
J Med Internet Res 2022;24(7):e27648
URL: https://www.jmir.org/2022/7/e27648
doi:10.2196/27648
PMID:35881456

Review

Mobile Apps for People With Rare Diseases: Review and Quality Assessment Using Mobile App Rating Scale

Sarah Hatem[1*], BSpHLSc, MPH; Janet C Long[1*], BSc Hons, MN, PhD; Stephanie Best[1,2], MSc, PhD; Zoe Fehlberg[1,2], BDes, MPH; Bróna Nic Giolla Easpaig[1], BSc Hons, PhD; Jeffrey Braithwaite[1], BA, DipLR, MIR Hons, MBA, PhD

[1]Australian Institute of Health Innovation, Macquarie University, North Ryde, Australia

[2]Australian Genomics, Murdoch Children's Research Institute, Melbourne, Australia

[*]these authors contributed equally

Corresponding Author:
Janet C Long, BSc Hons, MN, PhD
Australian Institute of Health Innovation
Macquarie University
Level 6, 75 Talavera Road
North Ryde, 2109
Australia
Phone: 61 0298502225
Email: janet.long@mq.edu.au

Abstract

Background: Mobile apps are becoming increasingly popular, with 5.70 million apps available in early 2021. Smartphones can provide portable and convenient access to health apps. Here, we consider apps for people with one of the estimated 7000 rare conditions, which are defined as having an incidence of <1 in 2000. The needs of people with rare conditions are known to be different from those of people with more common conditions. The former may be socially isolated (not knowing anyone else who has the condition) and may not be able to find reliable information about the disorder.

Objective: The aim of this review is to search for apps developed specifically for people diagnosed with a rare disease and to assess them for quality using the Mobile App Rating Scale (MARS). We examine features that address 6 identified needs of people with a rare disorder and make recommendations for future developers.

Methods: Google Play Store (Android) and Apple App Store (iOS) were searched for relevant health-related apps specifically for rare diseases. The search included the names of 10 rare disease groups. App quality was determined using MARS, assessing app engagement, functionality, aesthetics, and information.

Results: We found 29 relevant apps (from a total of 2272) addressing 14 rare diseases or disease groups. The most common rare conditions addressed were cystic fibrosis (n=6), hemophilia (n=5), and thalassemia (n=5). The most common app features were web-based information and symptom trackers. The mean MARS score was 3.44 (SD 0.84). Lowest scores were for engagement.

Conclusions: Most apps provided factual and visual information, providing tools for self-monitoring and resources to help improve interactions during health consultations. App origin and quality varied greatly. Developers are recommended to consider ways to make appropriate apps more easily identifiable to consumers, to always include high-quality information, improve engagement, provide qualitative evaluations of the app, and include consumers and clinicians in the design.

(J Med Internet Res 2022;24(7):e36691) doi:10.2196/36691

KEYWORDS

mobile apps; self-management; social support; rare disease; mobile phone

Introduction

Background

The number of smartphones is growing globally (6.259 million in 2021 and is expected to rise to an estimated 7.690 million by 2027 [1]). Downloadable apps are also growing exponentially, with a total of 5.70 million apps available in early 2021 from the 2 largest app stores, Google Play and the Apple App Store [2]. The ubiquity of mobile phones and their portability mean that apps to support health are convenient and accessible tools for large parts of the population [3]. Health apps can support healthy behaviors (eg, physical exercise [4], mindfulness [5]), support people with a specific condition (eg, diabetes [6], depression and anxiety [7]), or support health needs at certain life stages (eg, prenatal genetic testing [8], healthy aging [9]).

Rare diseases are a heterogenous group of conditions that are defined as having an incidence of <1 in 2000 [10,11]. An estimated 6000 to 8000 rare diseases have been discovered to date, affecting between 3.5% and 8% of the global population [11]. Rare diseases vary in origin and characteristics, including genetic conditions, infectious diseases, autoimmune diseases, and rare cancers [10,11]. Rare diseases are often harder to diagnose than commonly occurring diseases [12]. Although the importance of optimizing care for individuals with rare conditions is widely acknowledged, patients and their families report substantial barriers to accessing high-quality care and support following diagnosis [13]. Individuals living with a rare disease frequently report a lack of access to appropriate health care services, skilled health professionals, and management options [11]. Efforts to provide high-quality care are hampered by scarce research and medical knowledge (with sample sizes too low to run clinical trials for any individual disease), limited treatment options, and a lack of standardized guidelines for clinical management. For many rare diseases, health professionals may only see one case during their entire career [14].

Several studies have reported that people with rare conditions have differing needs from those with high-incidence conditions (eg, [15-17]). Although high-prevalence disorders may be no less distressing or onerous to care for, rare diseases have unique features: the lengthy odyssey to find a diagnosis, then to find appropriate specialists who know about the disorder; the lack of evidence about effective treatments, guidelines, or access to knowledgeable general health service providers; and isolation from peer support. Apps and other eHealth interventions (eg, telehealth, interactive websites, instant messaging, and web-based monitoring) may provide useful tools for health education, disease management, and patient advocacy in this cohort of patients but are likely to be different from apps designed for more common conditions.

This study complements a parallel study by the same authors (unpublished) looking at the needs of people with rare conditions that can be addressed by a range of eHealth tools including apps. That review found there to be 4 domains and 23 subdomains of the needs of people with a rare condition that could be addressed with eHealth interventions. The domains of need were support for self-management, access to high-quality information, access to appropriate specialist services, and social support.

App quality, especially from the viewpoint of consumers choosing an app from a vendor, has been explored and found to be lacking in a number of reviews of apps targeting high-incidence conditions (eg, [8,18,19]) and rare conditions [11]. Frequently mentioned are poor or absent reporting of trialing of the app or the evidence base on which the app is built [8] and difficult to understand information [11,20]. Moreover, the unique needs of people with rare diseases may mean that apps are not being designed to address those needs appropriately and acceptably. Reviews that identify unmet needs and highlight concerns regarding quality are therefore important.

Objectives

The aim of this review was to scope the nature and range of mobile apps developed specifically for patients diagnosed with a rare disease, or carers or parents of these patients. The Mobile App Rating Scale (MARS) [3] was used to assess the apps. The research questions are (1) What apps are available for people living with a rare disease? and (2) How do apps address reported needs and contribute to appropriate, high-quality care for people living with rare diseases?

To our knowledge, this is the first review of apps for people living with rare diseases. This review will focus on the specific needs of people with rare conditions, identifying the various benefits and shortcomings of existing apps, and will help inform the development of future apps for this often-overlooked group.

Methods

Search Strategy

The 2 most popular commercial app stores [2], Google Play Store (Android) and Apple App Store (iOS) were used to search for relevant apps. Searches were run in July 2021 using the Google Chrome feature, Google *incognito* mode, to lessen the influence of searchers' browser history. The 2 app stores were searched independently by 2 reviewers (SH and BNGE), where BNGE accessed the Google Play Store and SH accessed the Apple App Store using combinations of keywords: "Rare disease" AND "patient education" OR "health resource" OR "delivery of health care" OR "patient advocacy" OR "patient participation" OR "patient resource" OR "rare disease patient." This approach yielded few apps mostly because of the vagueness of the term "rare disease" (a term that covers 6000-8000 different conditions). The search was then modified to name 10 of the more common rare diseases or disease groups: cystic fibrosis, cystinosis, Fabry disease, hemophilia, hereditary angioedema, mitochondrial disease, narcolepsy, primary biliary cholangitis, spina bifida, and thalassemia (alpha and beta) [21].

App Selection

The inclusion criteria applied to the apps were (1) focus on a single rare disease or a group of rare diseases, with an incidence of <1 in 2000; (2) targeted at the patient or carer; (3) available in English; and (4) free to download. Apps were excluded if they (1) were directed solely at health professionals, (2) were solely collecting data from patients for research purposes, (3)

required a personalized access code or the user had to be in a specific geographic location to create an account, or (4) failed basic functional criteria such as download (after 2 attempts).

First, apps were screened by app name and basic description in the web-based Apple Store and Google Play Store by the authors SH and BNGE (who each performed 20% independently and then compared and discussed). Relevant apps were cataloged in a spreadsheet, and duplicates were removed by the same authors. Subsequently, the apps were formally screened on the web-based store's site by app name and description against the selection criteria by SH and ZF. Finally, the apps were downloaded on the personal phone devices of SH and ZF, and their features were examined to verify their inclusion. This was validated by the rest of the research team.

Data Extraction

For the included apps, the following information was extracted (following section 1 of the MARS tool described in the *Quality Approval* section below): app name, name of the rare disease or disease group, platform (iOS or Android), version, year of the latest update, language options, app developer, country of origin, age group of target audience, and purpose or aim. This information was identified from the *General description* features of each app. We further extracted details of app features, whether the app was consumer facing (for the person with the condition or their carer only) or collaborative (to be used with a health care professional) and whether the app required passive or active participant involvement to use (ie, passive requires no interaction or involvement from an individual [eg, general information] and active requires consumer interaction [eg, entering symptoms into a symptom tracker feature]).

Domains of needs of people with a rare disease were defined initially based on a separate review of the peer-reviewed literature that considered the use of all eHealth interventions, not just apps. The four domains of need found and examined were (1) social support, (2) access to high-quality information about their specific rare disease, (3) support for self-management, and (4) access to appropriate specialist services for their rare disease. As the data extraction proceeded for the apps in this review, this list was amended. Domain 4, access to appropriate specialist services for their rare disease, was deleted as not being present in the reviewed apps, and 2 additional domains were added: sharing patient data with the health team and contributing to a global research database or registry. This is shown in Table 1.

Following data extraction, the included apps were examined based on their features and classified according to their domains of need. This classification allowed us to compare the quality of apps with a similar purpose and determine whether apps were addressing the needs of people living with a rare disease.

Table 1. Domains of needs of people with a rare disease that may be addressed by apps.

Domain	Examples of app features
Domain 1: social support	Platform where people with the same rare disease can exchange experiences and information (may include input from a health care professional).
Domain 2: tools for improving consults with health professionals	Feature to prerecord questions and record the consultation, advice on setting goals, and so on.
Domain 3: high-quality information on rare disorders	Information on rare disease, negotiating with school or workplaces, up-to-date information on research and new clinical trials, guidelines, and links to appropriate websites.
Domain 4: self-management support	Symptom trackers, journals, medication reminders, appointment reminders, guidance for performing exercise or treatments (may include ability to share inputted data with health care professional).
Domain 5: improve coordination of care	Feature allows sharing of inputted data with multiple health care professionals.
Domain 6: contributing to a global research database or registry	Data can be entered by app user to contribute to global research.

Quality Appraisal

App quality was assessed using MARS, a tool specifically designed for rating and examining the quality of mobile apps used for health [3]. The tool consists of 6 categories. The first deals with classification (version, developer, targeted age group, etc). The second section rates the objective quality of the apps by assessing 4 attributes: *Engagement* (5 items), *Functionality* (4 items), *Aesthetics* (3 items), and *Information* (7 items). *Engagement* refers to whether the app is fun, interesting, customizable, or interactive (eg, it sends alerts, messages, reminders, enables sharing, and is well targeted to the audience). *Functionality* assesses app functioning, ease of use, navigation, flow logic, and gestural design of the app. *Aesthetics* assesses graphic design, overall visual appeal, color scheme, and stylistic consistency. *Information* assesses whether the apps contain high-quality information and references from a credible source. An example of the scoring criteria is given in Textbox 1 [3].

Scores for each section (3-7 criteria each) are computed as mean scores to allow for criteria that are not applicable (eg, criterion that ask about the quality of the information given in the app, but there is no information).

The last 2 sections of the MARS tool assess the subjective quality of the app and the information on the perceived impact of the app on the user. The authors' lack of familiarity with the conditions the apps addressed led to the decision to consider only objective criteria.

In total, 19 items were rated on a 5-point scale from 1 *Inadequate* to 5 *Excellent* and combined to create an overall objective quality score. Authors SH and ZF assessed the quality of each app, rated an initial 50% of the apps in parallel, and

confirmed acceptable interrater reliability. The 2 authors discussed and resolved any significant conflicts. Descriptive statistical analyses were performed using STATA (version SE 17; StataCorp).

Textbox 1. Scoring criteria for Aesthetics.

Scoring criteria

- Is the arrangement and size of buttons, icons, menus, or content on the screen appropriate or zoomable, if needed?

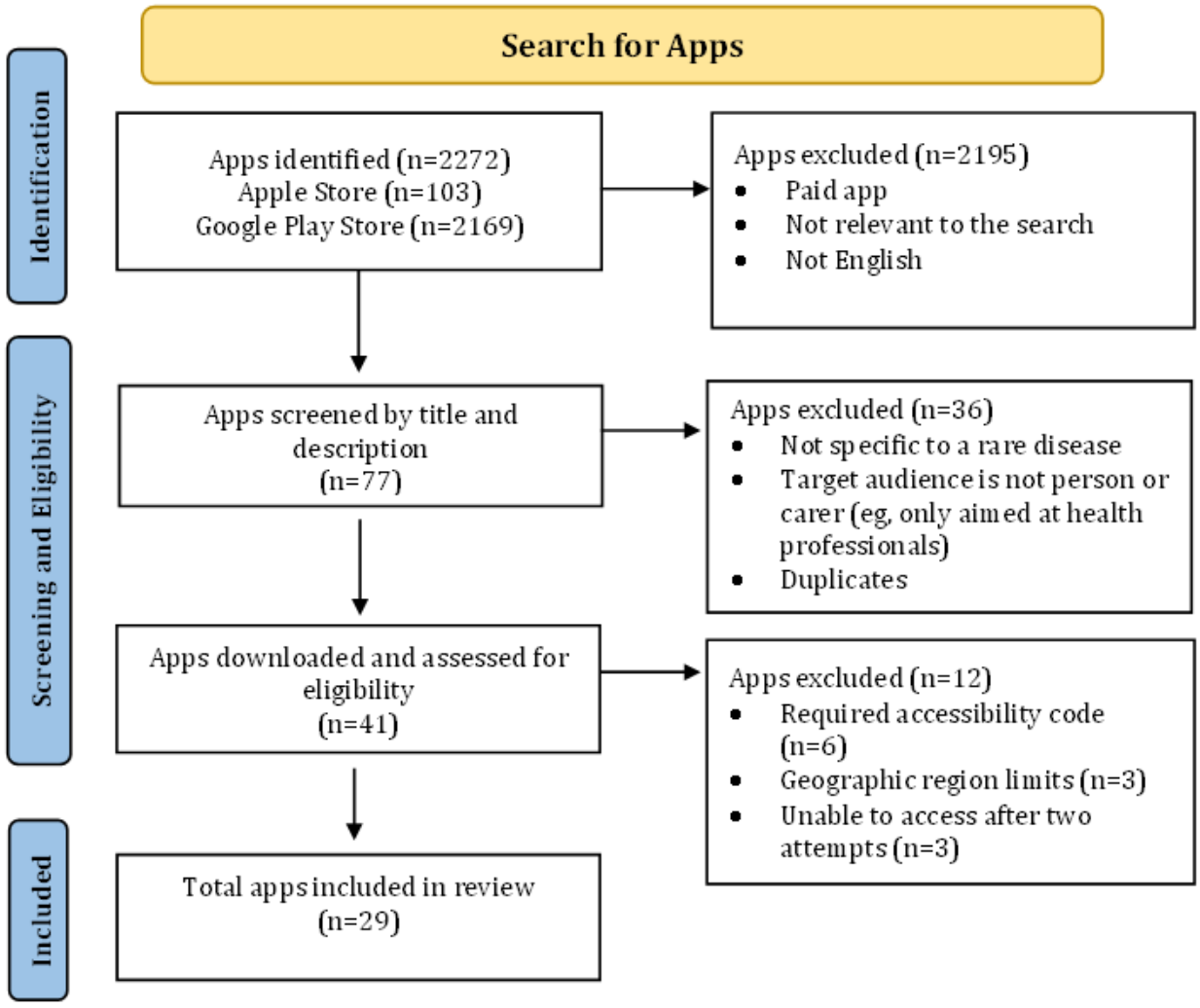
1. Very bad design, cluttered, some options impossible to select, locate, see, or read device display not optimized.
2. Bad design, random, unclear, some options difficult to select, locate, see, or read.
3. Satisfactory, few problems with selecting, locating, seeing, or reading items or with minor screen size problems.
4. Mostly clear, able to select, locate, see, or read items.
5. Professional, simple, clear, orderly, logically organized, device display optimized. Every design component has a purpose.

Results

Principal Results

Our search identified 2272 apps. Following the screening of titles and descriptions in the web-based store, a total of 96.61% (2195/2272) apps were excluded. Most of the excluded apps were not addressing a rare condition. The remaining 77 apps were recorded in a spreadsheet, with duplicates being removed (n=11) and screened again against the inclusion criteria. A total of 41 apps were then downloaded and evaluated against the selection criteria. In all, 29 remained for inclusion. The flow diagram (Figure 1 [22]) provides an overview of the selection process.

Figure 1. App selection process flowchart.



Characteristics of the Apps

Multimedia Appendix 1 summarizes the characteristics of the 29 included apps. Further details are provided in Multimedia Appendices 1 and 2. The apps addressed 14 different rare diseases or rare disease groups. Apps for people with cystic fibrosis were the most commonly identified (6/29, 21%). Our search did not locate any apps for Fabry disease, mitochondrial disease, and hereditary angioedema. Most apps (17/29, 59%) were available on both iOS and Android devices, and most (20/29, 69%) had been updated in the past 2 years. A total of 72% (21/29) of apps were only available in English, and 28% (8/29) were available in multiple languages. Most apps (10/29, 34%) did not specify their country of origin, and the most common developers were not-for-profit organizations or individuals (7/29, 24%). A total of 93% (27/29) of apps targeted consumers aged ≥12 years (2/29, 7% of apps for younger children), 55% (16/29) required active involvement by consumers, and 24% (7/29) were collaborative in nature.

Features of the Apps

The features of the apps are summarized below. Some apps included more than one feature.

In all, 52% (15/29) of apps included web-based information for patient education, including both factual and visual resources in the form of databases, downloadable information, informative videos, illustrations, and fact sheets. They described general disease information about the rare condition, disease history, signs and symptoms, management, and epidemiology where known. Exemplars are *Recognize Thalassemia Disease* and *Cystic Fibrosis: A Pocket Guide*. The *VASCERN app* contained the latest contact details for each relevant rare disease health care professional and patient organization, including services they provided, and hospital or clinic location information.

A total of 24% (7/29) of apps contained a feature that allowed them to export data trends they had recorded via symptom trackers and digital journals into a file that could then be shared with a health care professional during consultations (eg, *RareGuru* and *Cystic Fibrosis Manager)*. A total of 28% (8/29) of apps included a symptom tracker whereby individuals could record their condition-specific symptoms and side effects (eg, *Narcolepsy Monitor* and *Cystic Fibrosis Manager)*. A total of 24% (7/29) of apps involved the creation of a web-based profile where individuals could connect with others who were diagnosed with the same rare disease. Creating a profile allowed users to share resources, support each other, and stay updated about the latest news and research from different groups and foundations (eg, *RarePulse* and *RareGuru)*. A total of 24% (7/29) of apps included a digital journal, allowing users to record their experiences and feelings as their condition progressed, likely to be useful in helping individuals during consultations with health care professionals (eg, *PBC Health Storylines* and *Cystic Fibrosis Manager)*. Other features included medication reminders (5/29, 17%), medical appointment reminders (3/29, 10%; eg, *ThalTracker*), and the option to record medical tests (3/29, 10%). This last feature could be accessed by both health care professionals and individuals with the option to store the results in the patient profile (*PBC Health Storylines* and *Cystic Fibrosis Manager)*.

A total of 7% (2/29) of apps, *Breathe RM* and *PBC Health Storylines*, included the option to sync with a wearable device such as Fitbit or Apple Watch and allowed the app to access further information about vital signs and heart rate. The *Autogenic Drainage* app featured treatment support and *Cystic Fibrosis Downhill* featured an interactive game for children and adults as a form of patient education regarding their condition.

Quality Appraisal of the Apps

The total mean score of the 29 apps across the 4 MARS quality domains was, on a 5-point scale, 3.44 (SD 0.84). *Narcolepsy Monitor* had the maximum score at 4.69 and *Thalassemia Disease* had the minimum score at 1.95. The *Functionality* domain had the highest mean 4.23 (SD 0.62) across all apps and had the smallest variation in minimum and maximum scores (3.25 to 5). This means that navigation, ease of use, performance, and gestural design (eg, swipes and taps) across the apps were mostly intuitive and well designed. The lowest mean score was for the *Engagement* domain 2.94 (SD 1.08), which had the highest variation between minimum and maximum (1.2 to 4.8). For example, one of the lowest scoring apps for *Engagement* was *Autogenic Drainage*, which leads the user through a deep breathing exercise. Although it has some customizable features (duration of the session) and some encouraging messages (“Try not to cough!”), there are no graphics—only text and a basic timer—limiting its ability to engage.

The *Information* section of the MARS tool had the second lowest mean scores (second to *Engagement)*, reflecting deficits in quality, conciseness, ease of understanding, and use of the evidence base. One of the items in the information section of the MARS tool asks whether the “App has been trialed or tested and must be verified by evidence (in published scientific literature).” All the apps scored 0 for this item. Table 2 provides the mean scores across the 4 MARS quality domains and the total mean score, ranking from highest to lowest.

Table 2. App quality mean scores across the 4 Mobile App Rating Scale sections and total (out of 5).

	Engagement	Functionality	Aesthetics	Information	Total
Overall mean (SD; range)	2.94 (1.08; 1.2-4.8)	4.23 (0.62; 3.25-5)	3.40 (1.25; 1.33-5)	3.23 (0.98; 1.5-4.6)	3.45 (0.84; 1.95-4.69)
App name					
Narcolepsy Monitor	4.80	4.75	4.70	4.50	4.69
Cure SMA guide	4.60	5.00	4.70	4.40	4.68
Cystic Fibrosis Manager	4.60	4.75	4.70	4.00	4.51
Project Breathe or Breathe RM	3.80	4.75	5.00	4.00	4.39
RarePulse	4.40	4.75	3.67	4.60	4.36
PatientMpowerment for CF	3.80	5.00	4.67	3.75	4.30
RareGuru	3.8	4.50	4.33	4.00	4.16
MicroHealth Hemophilia	3.60	4.25	5.00	3.50	4.09
Cystinosis & me	4.00	4.50	3.67	4.00	4.04
Spina Bifida Association	2.40	5.00	4.67	3.83	3.98
HaemActive	2.60	4.75	5.00	3.50	3.96
PBC Health Storylines	3.80	4.00	4.00	3.33	3.78
ThalTracker	2.60	4.25	4.33	3.86	3.76
CANrecall	2.60	4.25	3.67	4.00	3.63
THALIA app	3.20	3.75	3.00	4.00	3.49
VASCERN app	2.20	4.00	3.67	4.00	3.47
ThaliMe	3.40	3.50	3.67	3.00	3.39
Cystic Fibrosis Downhill	3.40	4.00	3.67	2.00	3.27
Cystic Fibrosis: A Pocket Guide	2.40	3.75	3.33	3.17	3.16
PH Aware	2.80	3.33	2.33	4.00	3.12
Haemophilia Pal	2.60	4.00	3.00	2.50	3.03
MyHemophiliaTeam	3.20	3.25	2.67	2.75	2.97
Autogenic Drainage	2.60	4.00	2.00	2.50	2.78
Recognize Amyloidosis Disease	1.40	5.00	1.33	2.00	2.43
Recognize Thalassemia Disease	1.40	5.00	1.33	2.00	2.43
Easy Diagnosis- Thalassemia	1.20	5.00	1.33	1.60	2.28
Narcolepsy Disorder	1.40	3.25	1.67	1.67	2.00
Hemophilia Disease	1.40	3.25	1.67	1.67	2.00
Thalassemia Disease	1.40	3.25	1.67	1.50	1.95

Needs That the Apps Address

Overview

All 6 domains of need (defined in Table 1) were addressed by at least one of the 29 apps. More than half the apps (17/29, 59%) aimed to address domains 3 (access to high-quality information for their rare disease) and 4 (support for self-management). Less common (3/29, 10%) were apps that aimed to address domain 6 (contributing to a global research database or registry). This may have been owing to a majority of these apps being filtered out in the screening phase, as they were generally targeted at health care professionals and researchers instead of diagnosed individuals. Table 3 provides a summary of the apps and the needs they address. The number of needs each app aimed to address varied. Most apps were identified as addressing 1 or 2 needs (23/29, 79%). A total of 17% (5/29) of apps addressed 3 or 4 domains, and 3% (1/29) of app (*Cure SMA Guide*) aimed to address all 6 needs. In addition, 3% (1/29) of apps provided users with knowledge of how to negotiate or advocate for their needs in non–health care settings (schools, workplaces, or insurance).

Table 3. Domains of needs of people with a rare disease addressed by each app.

	Total domains, N=6	Need domains					
		Social support	Tools for improving consults with health professionals	High-quality information on rare disorders	Tools to support self-management	Sharing patient data with the health team or improving coordination of care or both	Facilitate contribution to a global research database or registry
Total	54	4	7	17	17	7	3
App name							
Narcolepsy Monitor	2		✓		✓	✓	
Cure SMA guide	6	✓	✓	✓	✓	✓	✓
Cystic Fibrosis Manager	4		✓	✓	✓	✓	
Project Breathe or Breathe RM	1				✓		
RarePulse	2			✓			✓
PatientMpowerment for CF	3				✓	✓	✓
RareGuru	2		✓		✓		
MicroHealth Hemophilia	2				✓	✓	
Cystinosis & me	2		✓		✓		
Spina Bifida Association	2			✓	✓		
HaemActive	2				✓		
PBC Health Storylines	2				✓	✓	
ThalTracker	2				✓	✓	
CANrecall	2		✓		✓		
THALIA app	3		✓	✓	✓		
VASCERN app	2	✓		✓			
ThaliMe	2	✓			✓		
Cystic Fibrosis Downhill	1			✓			
Cystic Fibrosis: A Pocket Guide	1			✓			
PH Aware	1			✓			
Haemophilia Pal	1				✓		
MyHemophiliaTeam	2	✓		✓			
Autogenic Drainage	1				✓		
Recognize Amyloidosis Disease	1			✓			
Recognize Thalassemia Disease	1			✓			
Easy Diagnosis- Thalassemia	1			✓			
Narcolepsy Disorder	1			✓			
Hemophilia Disease	1			✓			
Thalassemia Disease	1			✓			

Social Support

A total of 14% (4/29) of apps included characteristics that addressed social isolation. These apps allowed consumers to actively connect and liaise with others with the same condition, as well as others who were part of the general rare disease community. Features of these apps included the ability to create a web-based profile, connect on social media platforms, join community support groups, and take part in web-based forums and social events. These features were evident in *Cure SMA Guide*, the *VASCERN app*, *ThaliMe*, and *MyHemophiliaTeam*.

Improving Consults With Health Care Professionals

A total of 31% (9/29) of apps included tools that were targeted at improving communication between a patient and their health care professional or care team. These apps were often collaborative in nature, whereby patients could export their collected data to share with their care teams and health care professionals could connect with their patients via telehealth or direct messaging. Consumers had the opportunity to share real-time condition tracking with their health care professionals on some apps. Health care teams could digitally send and store laboratory results on the app and schedule appointments. These features were seen in *Narcolepsy Monitor, Cure SMA Guide, Cystic Fibrosis Manager, RareGuru, Cystinosis &me, CANrecall*, and *THALIA app*. An app, *CANrecall*, had a list of question prompt lists clinically designed to help individuals ask meaningful questions during a specialist appointment. This app also allowed patients to record their session and listen to their consultation later.

Access to High-Quality Information on Rare Disorders

A total of 59% (17/29) of apps included features that aimed to educate the user about a specified rare disease and provided access to a range of detailed information. This allowed the diagnosed individuals or their carers to expand their knowledge by reading reliable sources such as diagnosis, treatment, and management of the condition. Some apps offer these in an easy-to-read PDF format, video, illustrations, through external links, or questionnaires. This was displayed in *Cure SMA Guide, RarePulse, Narcolepsy Disorder, Cystic Fibrosis: A Pocket Guide, Spina Bifida Association*, and others (Table 3). An app, *Cystic Fibrosis Downhill*, creatively provided this using an interactive educational game. Others included the latest contact and location details of health care providers and patient organizations as well as the services they provided. The *Cure SMA Guide* included care guidelines, treatments, patient and hospitalization guidelines, and sample letters for emergency resources. These sample letters helped the individual liaise with hospitals, physicians, insurers, electricity billing companies, telephone providers, schools, and emergency departments.

Self-management Support

A total of 59% (17/29) of apps incorporated characteristics that enabled better self-management of the condition. Apps that met these characteristics commonly included symptom trackers, options for medication reminders and appointments, and export data trends for sharing with physicians and specialists. These apps ranged from patient facing only to collaborative, allowing access from both the health care team and the patient. Examples of these features were in *Narcolepsy Monitor*, *Cure SMA Guide*, and *Breathe RM*. Remote monitoring was also available in the *HaemActive* app, where patients could participate in exercises on their own or in consultation with their physiotherapist via a video function. Remote monitoring was also available in the *PatientMpowerment for CF* app.

Improve Coordination of Care

A total of 24% (7/29) of apps included features for improving the coordination of care between health care physicians and teams, allowing for collaborative participation. Members of the multidisciplinary team could log on to the app and update any important information such as previous medicine treatments and test results; for example, *Cystic Fibrosis Manager*.

Contributing to a Global Research Database or Registry

A total of 7% (2/29) of apps gave the consumer the opportunity to share their deidentified data to help contribute to global research for their condition, supporting the development of a new understanding of treatments for the disease. These were *Cure SMA Guide* and *PatientMpowerment for CF*. An app, *Rare Pulse*, provided features that helped patients and caregivers stay in the loop about the latest news and updates from different research groups and consumer advocacy agencies, while furnishing information on upcoming forums and events regarding the rare disease condition.

Discussion

Principal Findings

Apps have the potential to enhance the quality of care, management of care, and self-management for individuals diagnosed with a rare disease [23] and can contribute to patient empowerment in an area of health care that can be confusing to navigate. People with a rare disorder face unique barriers to accessing appropriate care, including social isolation, difficulty in comprehending health care practitioner communications, lack of information, complicated self-management, poor coordination of care between health teams, and obstacles in accessing relevant research. High-quality apps have the potential to provide cohesive and trustworthy information, tools to collect symptom and treatment data, and options to assist in liaising with health professionals during consultations.

This review found only a small number of health apps targeting rare conditions. There is the obvious problem that a low number of potential users is a poor incentive for developing an app, especially a commercial app (eg, in 2019, 28.7 million people in the United States had diabetes [24] compared with 35,000 people who had cystic fibrosis [25]). Apps for these rarer conditions, therefore, tend to be developed and funded by not-for-profit organizations—patient advocacy agencies or clinician-research collaborations. We note that 55% (16/29) of the apps found in this review were developed by not-for-profit patient organizations or research groups.

Searching for apps for people with a rare disorder highlighted 2 sets of issues. First, for us as researchers, using the term “rare disease” yielded few results. Adding selected named rare disease groups increased this yield; however, an exhaustive search was

not feasible. Rare disease group names are an inexact way to search. For example, there were no apps found for "mitochondrial disease." This rare disease group name covers over 350 different disorders, with names that use medical language or refer to the gene that is faulty (eg, DNM1L-related encephalopathy and Leber hereditary optic neuropathy). Moreover, many of these disorders go by several different names (eg, "MELAS" for ORPHA-550, OMIM 540000, mitochondrial encephalomyopathy lactic acidosis and stroke-like episodes or mitochondrial myopathy encephalopathy lactic acidosis and stroke-like episodes [26]). It was not feasible to consider a search for all these alternatives across all rare disorders.

Second, people with a rare disease may have difficulty finding relevant apps by searching by diagnosis. Precisely naming an app (eg, "My DNM1L-related encephalopathy") is an unlikely choice for developers, who are more likely to use a more appealing, colloquial name. Therefore, consumers may need to search for the disease group (eg, genetic disorders, neurological disorders, or brain disorders). This suggests that more thought should be given by developers to tag or label their apps so they can be easily found in consultation with clinical experts.

Many of the apps found in our review (16/29, 21%) provided some sort of web-based information about the disease, but as noted earlier, the *Information* items of the MARS received the second lowest mean scores of the 4 sections. Access to information from a reliable source is a requirement for people with rare conditions. Apps targeting people with rare diseases that do not contain information should seriously consider adding this function. The inclusion of reputable, high-quality, and easy-to-understand information and guidelines (where available) is recommended for all health apps [27]. For rare disease apps particularly, including links to internationally recognized rare disease information sites (eg, Orphanet) would be a useful addition.

We found no evidence in peer-reviewed literature that any app was formally tested for usability. Other reviews of apps for higher-prevalence conditions also found deficits in the information section of the MARS scale. For example, a review of apps for women undergoing prenatal testing found an absence of developmental testing with end users. In other items in the same review, in the information section, they found missing, incorrect, or difficult to understand information on the tests and a lack of visual information. We note that large-scale quantitative testing of the apps in our review may not be feasible because of the low number of users, but we suggest that qualitative evaluation is an acceptable alternative.

The impact of health apps may be constrained by their engagement with users. Loss of interest in using health apps over time is well documented (eg, [28]). The MARS tool scores for this review appeared similar in means and ranges across the respective sections of other reviews of health apps (eg, [3,29]). The lowest scoring section of the 4 MARS domains was for *Engagement*. This is a clear area in need of improvement. High-quality apps with simple functionality can encourage people with no prior experience using technology to embrace their use [29]. Interesting visuals and interactive functions are useful components for increasing users' desire to engage with an app [27,29].

Some apps incorporated active participation by users in the form of digital journals, medication reminders, appointment reminders, and capacity for recording test results. Symptom trackers (8/29, 28%) were the second most common feature of the apps and were noted to be more common than treatment facilitators (1/29, 3%). This likely reflects the lack of available treatment for many rare diseases. Self-assessments via smartphones can save time during a consultation and allow the patient to provide health care professionals with a more accurate update on their condition and improve their approach and confidence to engage in self-management practices [3].

Formal evaluation of clinical outcomes supported by the apps is desirable, but we found little evidence to support this. Only one app alluded to an evaluation in a clinical study, but the details could not be identified in PubMed or Google Scholar. Lack of clinical testing similarly reflects the lack of agreed clinical indicators, treatments, or sufficient numbers of patients to participate in trials in the field of rare diseases.

Strengths and Limitations

A strength of this paper includes the use of the high-quality appraisal tool, MARS [3]. The method used here to scope and assess apps is generalizable to other health conditions. Limitations were around the problems of searching for >7000 types of rare diseases. Results were limited by access to some apps; for example, ones that required an access code from an external source or required payment. The data security check could not be performed comprehensively with the available data.

Recommendations

This review suggests a number of recommendations for developers of apps for people with a rare disease: (1) appropriate apps that address rare conditions can be hard to find, so developers should carefully consider how to make their app easily identifiable in the app stores; (2) developers should consider the needs of people with rare conditions when developing apps and not just follow designs used by high-incidence conditions. In particular, the need for high-quality information and social support (ie, consumers may not know anyone else who has the condition) should be considered. Information should be sourced from high-quality sources and checked with clinicians who specialize in the disorder; (3) low scores found in this review for the MARS *Engagement* criteria argue for more thought being directed to designing interesting and engaging features; (4) the subjective star rating system is not always helpful for apps for people with a rare disease. Although the star system may be helpful for consumers trying to choose between 50 different apps for diabetes, most of the apps for rare diseases in this review had no rating or had only a handful of users each. Usability testing and other quality ratings of apps should be considered, as well as formal qualitative evaluations; and (5) following the expression "Nothing about us, without us" [30], the input of consumers via advocacy agencies (or as individuals) is important to develop a feasible, credible, and useful app.

Conclusions

This review aimed to identify and evaluate mobile apps in the Apple App Store and Google Play Store that addressed the needs of people diagnosed with a rare disease. Most apps focused on providing factual and visual information, tools for monitoring symptoms and resources to help improve interactions during health consultations. App quality and origin varied significantly. Developers are encouraged to consider the unique needs of people with a rare condition to make appropriate, engaging, easy-to-find, and useful apps for this often-neglected cohort.

Conflicts of Interest

None declared.

Multimedia Appendix 1
General characteristics of the included apps. Level of participant interaction involved in the app, for example, information only (passive) or more of a hands-on resource (active), for example, symptom trackers.
[PDF File (Adobe PDF File), 201 KB - jmir_v24i7e36691_app1.pdf]

Multimedia Appendix 2
Details of the included Apps.
[PDF File (Adobe PDF File), 143 KB - jmir_v24i7e36691_app2.pdf]

References

1. O'Dea S. Forecast number of mobile devices worldwide from 2020 to 2025 (in billions)*. Statista. 2021 Apr 01. URL: https://www.statista.com/statistics/245501/multiple-mobile-device-ownership-worldwide/ [accessed 2022-05-11]
2. Ceci L. Number of apps available in leading app stores 2022. Statista. 2022 Apr 27. URL: https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/ [accessed 2022-05-11]
3. 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]
4. Glynn LG, Hayes PS, Casey M, Glynn F, Alvarez-Iglesias A, Newell J, et al. Effectiveness of a smartphone application to promote physical activity in primary care: the SMART MOVE randomised controlled trial. Br J Gen Pract 2014 Jul;64(624):e384-e391 [FREE Full text] [doi: 10.3399/bjgp14X680461] [Medline: 24982490]
5. 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]
6. Eberle C, Löhnert M, Stichling S. Effectiveness of disease-specific mHealth apps in patients with diabetes mellitus: scoping review. JMIR Mhealth Uhealth 2021 Feb 15;9(2):e23477 [FREE Full text] [doi: 10.2196/23477] [Medline: 33587045]
7. Khademian F, Aslani A, Bastani P. The effects of mobile apps on stress, anxiety, and depression: overview of systematic reviews. Int J Technol Assess Health Care 2020 Dec 14;37:e4. [doi: 10.1017/S0266462320002093] [Medline: 33314997]
8. Wu K, Alegria R, Gonzalez J, Hu H, Wang H, Page R, et al. Characteristics and quality of mobile apps containing prenatal genetic testing information: systematic app store search and assessment. JMIR Mhealth Uhealth 2021 Oct 14;9(10):e30404 [FREE Full text] [doi: 10.2196/30404] [Medline: 34647898]
9. Helbostad JL, Vereijken B, Becker C, Todd C, Taraldsen K, Pijnappels M, et al. Mobile health applications to promote active and healthy ageing. Sensors (Basel) 2017 Mar 18;17(3):622 [FREE Full text] [doi: 10.3390/s17030622] [Medline: 28335475]
10. Baumbusch J, Mayer S, Sloan-Yip I. Alone in a crowd? Parents of children with rare diseases' experiences of navigating the healthcare system. J Genet Couns 2019 Feb;28(1):80-90. [doi: 10.1007/s10897-018-0294-9] [Medline: 30128673]
11. Anderson M, Elliott EJ, Zurynski YA. Australian families living with rare disease: experiences of diagnosis, health services use and needs for psychosocial support. Orphanet J Rare Dis 2013 Feb 11;8:22 [FREE Full text] [doi: 10.1186/1750-1172-8-22] [Medline: 23398775]
12. Willmen T, Völkel L, Ronicke S, Hirsch MC, Kaufeld J, Rychlik RP, et al. Health economic benefits through the use of diagnostic support systems and expert knowledge. BMC Health Serv Res 2021 Sep 09;21(1):947 [FREE Full text] [doi: 10.1186/s12913-021-06926-y] [Medline: 34503507]
13. Zurynski Y, Deverell M, Dalkeith T, Johnson S, Christodoulou J, Leonard H, APSU Rare Diseases Impacts on Families Study group. Australian children living with rare diseases: experiences of diagnosis and perceived consequences of diagnostic delays. Orphanet J Rare Dis 2017 Apr 11;12(1):68 [FREE Full text] [doi: 10.1186/s13023-017-0622-4] [Medline: 28399928]
14. Long JC, Best S, Hatem S, Theodorou T, Catton T, Murray S, et al. The long and winding road: perspectives of people and parents of children with mitochondrial conditions negotiating management after diagnosis. Orphanet J Rare Dis 2021 Jul 13;16(1):310 [FREE Full text] [doi: 10.1186/s13023-021-01939-6] [Medline: 34256797]

15. Depping MK, Uhlenbusch N, von Kodolitsch Y, Klose HF, Mautner V, Löwe B. Supportive care needs of patients with rare chronic diseases: multi-method, cross-sectional study. Orphanet J Rare Dis 2021 Jan 22;16(1):44 [FREE Full text] [doi: 10.1186/s13023-020-01660-w] [Medline: 33482869]
16. Kreuter M, Bendstrup E, Russell A, Bajwah S, Lindell K, Adir Y, et al. Palliative care in interstitial lung disease: living well. Lancet Respir Med 2017 Dec;5(12):968-980. [doi: 10.1016/S2213-2600(17)30383-1] [Medline: 29033267]
17. Quinn L, Davis K, Yee A, Snyder H. Understanding genetic learning needs of people affected by rare disease. J Genet Couns 2020 Dec;29(6):1050-1058. [doi: 10.1002/jgc4.1233] [Medline: 32128950]
18. McKay FH, Cheng C, Wright A, Shill J, Stephens H, Uccellini M. Evaluating mobile phone applications for health behaviour change: a systematic review. J Telemed Telecare 2018 Jan;24(1):22-30. [doi: 10.1177/1357633X16673538] [Medline: 27760883]
19. Sneha S, Thalla S, Rischie I, Shahriar H. Health Internet technology for chronic conditions: review of diabetes management apps. JMIR Diabetes 2021 Aug 31;6(3):e17431 [FREE Full text] [doi: 10.2196/17431] [Medline: 34463627]
20. Heilman CB. Readability of patient education materials. World Neurosurg 2013 Nov;80(5):e109-e110. [doi: 10.1016/j.wneu.2011.11.013] [Medline: 22381326]
21. Ghosh I. Infographic: which rare diseases are the most common? Visual Capitalist. 2019. URL: https://www.visualcapitalist.com/which-rare-diseases-are-the-most-common/ [accessed 2022-05-11]
22. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, 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]
23. Bakogiannis C, Tsarouchas A, Mouselimis D, Lazaridis C, Theofillogianakos EK, Billis A, et al. A patient-oriented app (ThessHF) to improve self-care quality in heart failure: from evidence-based design to pilot study. JMIR Mhealth Uhealth 2021 Apr 13;9(4):e24271 [FREE Full text] [doi: 10.2196/24271] [Medline: 33847599]
24. Prevalence of diagnosed diabetes. Centers for Disease Control and Prevention. 2022. URL: https://www.cdc.gov/diabetes/data/statistics-report/diagnosed-diabetes.html [accessed 2022-05-18]
25. Cystic fibrosis. Centers for Disease Control and Prevention. 2022. URL: https://www.cdc.gov/genomics/disease/cystic_fibrosis.htm [accessed 2022-05-18]
26. Search for a rare disease: mitochondrial disease. Orphanet. 2022. URL: https://www.orpha.net/consor/cgi-bin/Disease_Search_Simple.php?lng=EN [accessed 2022-06-23]
27. Knitza J, Tascilar K, Messner E, Meyer M, Vossen D, Pulla A, et al. German mobile apps in rheumatology: review and analysis using the Mobile Application Rating Scale (MARS). JMIR Mhealth Uhealth 2019 Aug 05;7(8):e14991 [FREE Full text] [doi: 10.2196/14991] [Medline: 31381501]
28. Owen JE, Jaworski BK, Kuhn E, Makin-Byrd KN, Ramsey KM, Hoffman JE. mHealth in the wild: using novel data to examine the reach, use, and impact of PTSD coach. JMIR Ment Health 2015 Mar 25;2(1):e7 [FREE Full text] [doi: 10.2196/mental.3935] [Medline: 26543913]
29. Mandracchia F, Llauradó E, Tarro L, Valls RM, Solà R. Mobile phone apps for food allergies or intolerances in app stores: systematic search and quality assessment using the mobile app rating scale (MARS). JMIR Mhealth Uhealth 2020 Sep 16;8(9):e18339 [FREE Full text] [doi: 10.2196/18339] [Medline: 32936078]
30. Charlton J. Nothing About Us Without Us: Disability Oppression and Empowerment. Berkeley, CA, USA: University of California Press; 1998.

Abbreviations

MARS: Mobile App Rating Scale

Edited by A Mavragani; submitted 07.02.22; peer-reviewed by K Wendrich, S Rostam Niakan Kalhori, A Teles; comments to author 09.04.22; revised version received 27.05.22; accepted 13.06.22; published 26.07.22.

Please cite as:
Hatem S, Long JC, Best S, Fehlberg Z, Nic Giolla Easpaig B, Braithwaite J
Mobile Apps for People With Rare Diseases: Review and Quality Assessment Using Mobile App Rating Scale
J Med Internet Res 2022;24(7):e36691
URL: https://www.jmir.org/2022/7/e36691
doi:10.2196/36691
PMID:35881435

Review

A Comprehensive Literature Search of Digital Health Technology Use in Neurological Conditions: Review of Digital Tools to Promote Self-management and Support

John Henry Spreadbury[1,2], BSc, MSc, PhD; Alex Young[2], MSc, PhD; Christopher Myles Kipps[2,3], MBBS, PhD

[1]Faculty of Medicine, University of Southampton, University Hospital Southampton NHS Foundation Trust, Southampton, United Kingdom
[2]Clinical Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, United Kingdom
[3]Wessex Neurological Centre, University Hospital Southampton NHS Foundation Trust, Southampton, United Kingdom

Corresponding Author:
John Henry Spreadbury, BSc, MSc, PhD
Faculty of Medicine
University of Southampton
University Hospital Southampton NHS Foundation Trust
Tremona Road
Southampton, SO16 6YD
United Kingdom
Phone: 44 02381206132 ext 6132
Email: jhs101@soton.ac.uk

Abstract

Background: The use of digital health technology to promote and deliver postdiagnostic care in neurological conditions is becoming increasingly common. However, the range of digital tools available across different neurological conditions and how they facilitate self-management are unclear.

Objective: This review aims to identify digital tools that promote self-management in neurological conditions and to investigate their underlying functionality and salient clinical outcomes.

Methods: We conducted a search of 6 databases (ie, CINAHL, EMBASE, MEDLINE, PsycINFO, Web of Science, and the Cochrane Review) using free text and equivalent database-controlled vocabulary terms.

Results: We identified 27 published articles reporting 17 self-management digital tools. Multiple sclerosis (MS) had the highest number of digital tools followed by epilepsy, stroke, and headache and migraine with a similar number, and then pain. The majority were aimed at patients with a minority for carers. There were 5 broad categories of functionality promoting self-management: (1) knowledge and understanding; (2) behavior modification; (3) self-management support; (4) facilitating communication; and (5) recording condition characteristics. Salient clinical outcomes included improvements in self-management, self-efficacy, coping, depression, and fatigue.

Conclusions: There now exist numerous digital tools to support user self-management, yet relatively few are described in the literature. More research is needed to investigate their use, effectiveness, and sustainability, as well as how this interacts with increasing disability, and their integration within formal neurological care environments.

(J Med Internet Res 2022;24(7):e31929) doi:10.2196/31929

KEYWORDS

digital health technology; digital tools; neurology; patients; self-management

Introduction

Background

Neurological conditions present a human and economic challenge worldwide. How best to manage them remains a perennial issue. Digital health technology offers a potential solution. It would seem plausible that digital technology could play some role in supporting patients in self-management or health care professionals in the delivery of care. However, the digital health market contains a bewildering variety of websites, online platforms, and apps, some with empirical support, making it difficult to make sense of what is available, and their potential benefits. The objective of this paper was to conduct a literature

search of the research on digital health technology in the self-management of neurological conditions, and to investigate what functions the technology provides and what benefits to users have been reported.

Neurological Conditions

Neurological conditions refer to a group of medical disorders often resulting from disease or physical damage that affect the brain, or central or peripheral nervous systems. They can negatively impact patient mental health [1-4], psychological well-being [3], life satisfaction [3], health-related quality of life [5-8], cognitive functioning [3], and social support [9]. Worldwide, they are identified as significant predictors of disability and death [10,11]. They can also be detrimental to caregivers in terms of their mental health, quality of life, and caregiver burden [12-14].

As well as a human burden there is also an economic one. In the United Kingdom, statistics from the Neurological Alliance [15] indicate 16.5 million people in England have a neurological disorder. This statistic is equivalent to 1 in 6 of the population, and a prevalence believed to be increasing [15]. It is estimated that the National Health Service (NHS) cost of addressing neurological disorders is around £4.4 (US $3.0) billion [15], and may account for up to 14% of social care spending [5].

Many neurological conditions will be long term and incurable, and have symptoms that produce persistent or sporadic difficulties. Their onset may be sudden or gradual and their trajectories are marked by variance in stability or progression. Treatment and management may vary in complexity and include a combination of medication, rehabilitation, information, and support, and the involvement of a range of health care, allied health, and social care professionals [5].

Neurological conditions are generally managed in the community and there is increasing recognition of the importance of individuals self-managing their conditions [5,16]. Recent qualitative research by Kilinc et al [16] demonstrated the complex psychological and behavioral processes underlying self-management in neurological patients. The involvement of technology is one approach to supporting such efforts [5], while research by Gandy et al [3] indicated that there is interest among patients in web-based platforms to promote self-management.

Digital Health Technology

Digital health technology, including terms such as eHealth, mobile health (mHealth), and digital tools, refers to the utilization, or application, of internet and smart-based technology to the promotion of health or health care [17]. Innovative technologies such as wearable devices, smartphone apps, internet-based self-help platforms, and health record databases have the ability to record, store, or present health-related data. This information can then be used to enhance the understanding, management, or monitoring of medical conditions by patients, carers, or health care professionals.

A range of digital technologies have already been applied to several individual neurological conditions such as epilepsy [18], MS [19], headache and migraine [20], Parkinson disease [21], and acquired brain injury [22]. There appears use and interest, at least in the short-term, and some evidence, with regard to web-based platforms, of a potential beneficial influence on mental health and quality of life [23]. However, it remains unclear how digital technologies become normalized within health behaviors and systems of care delivery in the medium-to-longer term. Furthermore, there may be significant patient and care-provider barriers that need to be considered [18,24,25].

A limitation of the present literature is that recent reviews and commentaries have tended to focus on individual neurological conditions (eg, [18-21,26-30]). There is an absence of reviews presenting digital tools across conditions that makes it difficult for clinicians and researchers, especially those new to digital health, to make comparisons, evaluations, and recommendations. It would be advantageous to know what digital tools are available to different patient groups, the underlying functionalities that support or promote self-management, and any salient psychosocial or clinical benefits for users identified.

Aims

The literature search had several interrelated aims. First, we aimed to obtain an overview of the research on the use of digital health technology in the self-management of neurological conditions. Second, we aimed to identify the different types of digital health tools used by patients, carers, and health care professionals. Third, we aimed to develop an understanding of the underlying functionalities that allow digital health tools to support or promote self-management. Finally, we aimed to identify any salient outcomes, in terms of psychosocial or clinical benefits for users, associated with digital health technology use.

Methods

Literature Search Databases and Search Terms

We conducted a search of 6 databases: CINAHL, EMBASE, MEDLINE, PsycINFO, Web of Science, and the Cochrane Review. The searches were conducted using free text and equivalent database-controlled vocabulary terms. Search terms used were iteratively generated and informed by our interest in investigating digital health technology use in neurological conditions and neurodegenerative diseases. Multimedia Appendix 1 provides an example of the search terms.

Within each database, search terms were grouped into 2 categories: condition terms (ie, neurological conditions and neurodegenerative diseases) and digital technology terms. Search terms were combined using standard AND/OR commands. Where possible, filters were applied within databases to restrict searches to human participants, adults, and beginning from January 2000 onward. Searches spanned from January 2000 to February 2020, and were rerun in January 2021.

Inclusion Criteria

The inclusion criteria for articles were as follows: Research conducted with human participants and published in English. Studies that had a focus on the use of digital health technology to help support self-management in patients or caregivers living with a neurological condition or neurodegenerative disease.

Self-management was understood to refer to activities used to control a medical condition or maintain optimal health [31]. The self-management health component had to be delivered digitally, for example, via a computer, mobile/tablet app, or over the internet.

Exclusion Criteria

Articles were excluded if they were not conducted with human participants or if they focused on artificial intelligence, biochemistry, computational modeling, diagnosis/assessment, cognitive stimulation/training, epidemiology, genetics, neuroimaging, neuropathology, physiotherapy, rehabilitation, scale development/validation, sensor technology, treatment, or interventions delivered by telephone. These areas were excluded to help narrow down the focus of digital health technology involved in self-management. Literature reviews, book chapters, study protocols, conference presentations, poster presentations, and unpublished theses were excluded.

Search Methodology

Multimedia Appendix 2 shows that the overall search resulted in 26,572 articles being identified. Articles were downloaded into an Endnote library and duplicates were removed. The remaining articles were then exported to Rayyan reference management software, which allowed for the collaborative screening of articles by 2 reviewers. Articles were screened by reading the title and abstract of each article and applying the inclusion and exclusion criteria. Any articles where reviewers had conflicting opinions were discussed at the end of this process until consensus was met on inclusion or exclusion for full-text screening.

Following title and abstract screening, 96 articles moved forward to full-text screening. Microsoft Excel was used to list the 96 articles and extract salient information related to each article's aims, methodology, results, and use of digital health technology. Full-text screening resulted in 45 articles being excluded for not meeting the inclusion criteria.

Rerunning the database searches and using keyword searches in Google Scholar resulted in 2 further articles being included. This resulted in a final total of 53 articles. A total of 27 articles focused on digital health technology use in neurological conditions and 26 on digital health technology use in dementia. The present paper only discusses the neurological condition articles.

Study Methodological Quality and Value of Findings

We used the Critical Appraisal Skills Programme (CASP) Appraisal Tool to evaluate the methodological quality and value of the findings reported for each of the 27 included articles. All of the articles were considered to be of satisfactory methodological quality and produced findings of value. No cut-off scores were used and no articles were excluded as a result of using the tool.

Analysis

The analysis is reported in 4 parts. First, we describe the contextual background of the articles. Using a data extraction table, we extracted from each article information about its nationality, the type of neurological condition studied, and methodological details (eg, participants, designs, outcome measures).

Second, we describe the digital health tools identified. From each article, we extracted information about the digital tool reported, including its name, the neurological condition it addressed, the format of the technology, its users, and its broad aims.

Third, we describe the underlying functionalities of the digital health tools that appeared to promote or support self-management. This information was obtained by extracting from each article the description of how each digital tool functioned. By iteratively reading through descriptions several different categories of function could be identified across the articles. These categories were then grouped together based on the similarity of functions to create 5 overarching categories that represented the main functionalities provided by the digital tools.

Finally, we describe salient psychosocial and clinical benefits associated with the digital health tools. This information was obtained by extracting the main outcomes reported that reflected psychosocial or clinical benefits to users.

Preliminary data analysis, findings, and interpretations from the review were presented at internal research group meetings for sense checking and feedback.

Results

Contextual Background

The search identified 27 articles. These articles came from 9 different countries. The majority of articles were from the United States with 15. This was followed by 4 articles from Holland and 2 from Australia. There was 1 article each from Belgium, Germany, New Zealand, and Turkey. One additional article reported on a sample including participants from the UK and Canada, and 1 with participants from the UK and New Zealand.

A total of 10 articles focused on MS, 6 on epilepsy, 6 on stroke, 4 on headache or migraine, and 1 on pain. Two articles with a focus on MS also included participants with Parkinson disease and postpolio syndrome. The majority of articles centered on patients (n=21), with a minority on carers (n=4). One article included patients and carers, and 1 patients and health care professionals.

The majority of articles reported studies using quantitative or mixed quantitative-qualitative designs. Only 2 articles reported qualitative studies. Across the articles a range of measurements were employed, including widely used questionnaire instruments (eg, on mental health, fatigue), process evaluation metrics (eg, usability, satisfaction), digital technology system metrics or stored data (eg, recorded usage of a digital tool), and open-ended questions (eg, on subjective experience).

We identified approximately 100 questionnaire instruments, including instruments used more than once. When these instruments were broadly grouped together based on the similarity of construct being measured, 16 measurement domains could be identified (Table 1). Among the most prevalent areas

measured were mental health, quality of life, fatigue/physical activity, disability, and self-efficacy.

Table 1. Estimate of measurement domains by percentage of questionnaire instruments used.

Measurement domain	%
Mental health	16
Quality of life/life satisfaction	13
Fatigue/activity	10
Disability	10
Self-efficacy	9
Coping/control	9
Self-management	6
Stress	4
Usability	4
Sleep quality	3
Social support	3
Care satisfaction/quality	3
Health care utilization	3
Improvement	2
Health status	2
Condition knowledge	2
Other	3

Digital Tools and Aims

Table 2 shows that 17 different digital tools were reported across the articles. A number of them, for example, PatientsLikeMe, WebEase, Mymigraine, and Caring-Web, were reported by more than 1 article. The majority of digital tools were website/web-based platforms and a minority were smartphone apps.

MS had the highest number of reported digital tools with 8, and this was followed by epilepsy and stroke both with 3, and headache and migraine with 2. The platform painACTION was reported in 2 different conditions—headache and migraine, and pain. The majority of digital tools focused on patients, while only 2 platforms, both related to stroke, focused on carers.

In the MS group, there were tools that specifically targeted fatigue and depression as well as personal health record management. In epilepsy, there were tools that involved collaborative self-management with a health care professional and information sharing within a health-related social network. For stroke, provision of stroke-related education was offered to carers and patients. In headache and migraine, tools provided training to promote self-management potential, and in pain there was a digital tool that addressed cognitive and emotional aspects of pain self-management.

Table 2. Digital health technology by neurological condition, type of technology, users, and aim.

Condition and digital technology name	Type of technology	Users	Broad aim
Multiple sclerosis			
MS Energize	Smartphone app	Patients	Fatigue self-management
Problem Solving Therapy	Website/web-based platform	Patients	Depression self-management
MS TeleCoach	Smartphone app	Patients	Physical activity/fatigue self-management
Deprexis	Website/web-based platform	Patients	Treatment of depression
MSdialog	Web-based/smartphone app	Patients	Multiple sclerosis management/health data sharing
Mellen Center Care Online	Website/web-based platform	Patients	Personal health (record) management/self-management
PatientSite	Website/web-based platform	Patients/health care professionals	Personal health (record) management system
MSInvigor8	Website/web-based platform	Patients	Self-management/fatigue self-management
Epilepsy			
MINDSET	Tablet-based platform	Patients/health care professionals	Shared clinical decision tool/self-management
PatientsLikeMe[a]	Website/web-based platform	Patients	Data sharing/health social network/understanding
WebEase[a]	Website/web-based platform	Patients	Epilepsy self-management
Stroke			
Stroke Carer Support	Website/web-based platform	Carers	Carer education/enhance understanding/capability
Caring Web[a]	Website/web-based platform	Carers	Carer education/support
Post-Discharge Support	Website/web-based platform	Patients	Education/information provision/coping
Headache/migraine			
painACTION[a]	Website/web-based platform	Patients	Migraine self-management/coping/self-efficacy
Mymigraine[a]	Website/web-based platform	Patients	Behavior training/self-management
Pain			
painACTION[a]	Website/web-based platform	Patients	Pain self-management
Multiple sclerosis, Parkinson disease, postpolio syndrome			
Fatigue Self-Management Program[a]	Website/web-based platform	Patients	Fatigue self-management

[a]Reported in more than 1 article.

Digital Tools and Functionality

We identified 5 broad categories of interrelated functionality across digital tools: (1) knowledge and understanding; (2) behavior modification; (3) self-management support; (4) facilitating communication; and (5) recording condition characteristics.

Knowledge and Understanding

Around two-thirds of the digital tools had functionality involving increasing neurological condition knowledge and understanding. This category included tools providing psychoeducational/self-help information and cognitive behavior therapy guidance. Users could engage with learning-orientated "modules" or "lessons," often presented using interactive multimedia formats, and in some cases the completion of "homework" activities [32-38].

Around half of the digital tools provided some form of psychoeducational/self-help information. This support could include information on medical or psychosocial issues, coping and managing, or healthy living, and in some cases internet links to related resources [32,33,35,38-41]. In the case of stroke carers, there was comprehensive information on caring for a patient with stroke at home [41,42].

Approximately one-third of digital tools drew on or included a cognitive behavior therapy component. This function involved engagement with learning activities that encouraged users to address challenging condition–related cognitions, behaviors, lifestyles, or expectations; increase self-awareness or

self-understanding; and learn new skills and their application [23,32-34,36,38,43,44].

Behavior Modification

Around one-third of digital tools aimed to prompt behavior modification and included a focus on stimulating behavior change and providing coaching or motivation. A small number of tools addressed behavior change using activities such as assessment and evaluation of behavior, establishing behavior objectives, and utilizing "action plans" [39,45-47]. Selected digital tools also had the ability to provide user feedback, "motivational" messaging, advice, reminders, or encouragement [35,39,48-50].

Self-management Support

Overlapping with behavior modification were digital tools with the function of facilitating users in psychological or tangible self-management. This function assisted users in contemplating their own or preferable self-management, in some cases bolstered by feedback, and encouraged consideration of processes or targets to aid enhancement [39,45-47,49]. Tangible self-management was offered by the PatientSite platform that permitted users to access aspects of their own health record including their medical record, test results, health care appointments, and medication prescriptions [40].

Facilitating Communication

Approximately half of the digital tools facilitated communication either between users and health care professionals or peer-to-peer. Communication was often asynchronous, could be condition or intervention related, and used various formats, for example, email or discussion groups [38,39,44,46]. User communication with health care professionals could involve sharing health information, making requests, or asking questions [40,41,51], while health care professional communication could take the form of replies to users, supportive messages, reminders, or feedback [35,38,44]. Peer-to-peer communication could involve sharing experiences or advice [7,35,39].

Recording Condition Characteristics

Around one-third of digital tools included a function for recording condition-related information that could then be "tracked," "monitored," or "shared" to enhance management or understanding [7,34,39,45,51,52]. Finally, there was a digital tool, Caring Web, that included an entertainment function, whereby users had access to amusements (eg, "jokes" and "games") and topical news features [41].

Digital Tools and Outcomes

For the majority of digital tools some form of acceptability (eg, effectiveness, feasibility) was reported. This could be in the context of user responses, as a method of data collection, or in producing certain outcomes.

Self-management per se was seldom measured but instead proxies were used such as self-efficacy or coping. Where condition self-management could be directly measured as in epilepsy, digital tools such as WebEase and PatientsLikeMe were associated with enhanced self-management [39,52]. Across the conditions migraine, epilepsy, and a sample including MS, Parkinson disease, and postpolio syndrome, the digital tools painACTION, WebEase, Fatigue Self-Management, PatientsLikeMe, and Mymigraine were associated with improved condition-related self-efficacy [33,35,36,39,44,46,52]. Across the conditions migraine, pain, and stroke, the digital tools Mymigraine, painACTION, and Post-Discharge Support, respectively, were associated with either increased coping or use of positive coping strategies [33,34,37,50].

Depression was a frequently measured outcome and produced mixed findings. Scales used to measure depression included the Beck Depression Inventory [53]; Depression, Anxiety, and Stress Scale [54]; Hospital Anxiety and Depression Scale [55]; and the Centre for Epidemiological Studies Depression scale [56]. Across the conditions MS, migraine, and pain, digital tools such as Problem Solving Therapy, painACTION, Deprexis, and Fatigue Self-Management were associated with lower depression [23,33-35,43]. However, across the conditions MS and stroke, digital tools such as MS TeleCoach, Fatigue Self-Management, MSInvigor8, and Caring-Web showed no association with depression [38,48,57,58].

An outcome frequently measured in MS articles was fatigue and robust findings were identified. Measures of fatigue included the Fatigue Scale for Motor and Cognitive Functions [59] and a version of the Fatigue Impact Scale [60]. Digital tools such as MS TeleCoach, Deprexis, Fatigue Self-Management, and MSInvigor8 were associated with better fatigue scores [23,35,38,48,57]. Although quality of life was frequently measured, only the digital tool Deprexis appeared to show a positive influence [23].

Discussion

Principal Findings

This review provides an overview of self-management digital tools across a number of neurological conditions. The findings offer a complementary perspective to the literature on digital tool development and implementation by focusing on functionality and beneficial outcomes. Five broad categories of interrelated functions can be discerned that allow digital tools to promote self-management. Among these functions are the provision of information to increase knowledge and understanding; encouragement of positive behavior change; support in psychological and tangible self-management; facilitating communication between users and health care professionals or users in a similar situation; and the ability to record, monitor, and share condition information.

The digital tools appeared modestly associated with psychosocial or clinical benefits to users. Depression was frequently measured and yet while some digital tools indicated potential for reducing depression, for others there was no association. By contrast, a number of MS digital tools demonstrated some potential in managing fatigue. Interestingly, self-management in itself was seldom measured outside of epilepsy; however, certain digital tools were associated with increased self-efficacy and use of positive coping strategies.

Across the literature we found little discussion about health service adoption or endorsement of digital tools or how they fit

with the formal neurological care individuals receive [32,35]. For health service adoption, functionalities and user outcomes should be compatible with existing models of care. Functionalities such as promoting knowledge and understanding, facilitating communication with health care professionals, and recording condition information may lend themselves well to health service adoption. However, the evidence of user benefits may still be too limited. Indeed, future research should test digital tools by embedding and evaluating them within clinical care pathways. As such, the digital tools reviewed may best be considered as supplementary resources to any formal neurological care being received.

There was also little discussion across the literature about uptake and continued use of digital tools beyond a research context [38,39]. As part of analyzing articles, using the internet to conduct searches, we found it difficult to identify whether some digital tools were still in use or not. Indeed, future research could attempt to establish how many of the digital tools reported are still in use and how many have been abandoned and why (eg, changes in technology, low user uptake, cost).

There are a number of methodological limitations that should be considered. We excluded articles focused on assessment, cognitive training, physiotherapy, and sensor technology and this could have influenced the findings. These articles were excluded as at an early stage of screening it was judged that these areas contributed more to diagnosis, rehabilitation, and assistive technology than self-management. We did not identify as many self-management apps as we had expected; this may have been caused by not including within our searches the brand names of any apps or app marketplaces; however, more likely, many apps exist that are simply not reported in the scientific literature. Furthermore, we did not search the gray literature for self-management apps.

Future research should try to establish user preferences toward identifying the functions used most frequently, considered most useful, and that produce clinical benefits. Research should also consider whether user needs and preferences are being addressed. Prospective research could investigate the effect of medium-to-longer-term usage on user outcomes, and the effect on formal neurological care usage. Understanding the effect of integrating data from digital tools into formal clinical records, and the impact of utilizing multiple different tools simultaneously would also be worthwhile.

Conclusions

Digital health technology has been applied to a number of neurological conditions, yet there is a relatively limited literature on its use and usefulness in the context of self-management. It is likely that numerous other apps and websites have yet to enter the research literature. Detailed analysis and description of the self-management process is lacking as are condition-specific self-management scales, comparison of digital tools, and consideration of comparative outcomes. There appear to be modest associations with psychosocial or clinical outcomes but evaluation is needed of whether certain functionalities predict certain outcomes.

Acknowledgments

This report is independent research funded by the National Institute for Health and Care Research ARC Wessex. The views expressed in this publication are those of the author(s) and not necessarily those of the National Institute for Health Research or the Department of Health and Social Care.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Example of search terms.
[DOCX File , 15 KB - jmir_v24i7e31929_app1.docx]

Multimedia Appendix 2
Flowchart of the literature search.
[PNG File , 24 KB - jmir_v24i7e31929_app2.png]

References

1. Bulloch AG, Fiest KM, Williams JV, Lavorato DH, Berzins SA, Jetté N, et al. Depression--a common disorder across a broad spectrum of neurological conditions: a cross-sectional nationally representative survey. Gen Hosp Psychiatry 2015 Nov;37(6):507-512. [doi: 10.1016/j.genhosppsych.2015.06.007] [Medline: 26153456]
2. Gandy M, Karin E, McDonald S, Meares S, Scott AJ, Titov N, et al. A feasibility trial of an internet-delivered psychological intervention to manage mental health and functional outcomes in neurological disorders. J Psychosom Res 2020 Sep;136:110173. [doi: 10.1016/j.jpsychores.2020.110173] [Medline: 32623193]
3. Gandy M, Karin E, Fogliati VJ, Meares S, Nielssen O, Titov N, et al. Emotional and cognitive difficulties, help-seeking, and barriers to treatment in neurological disorders. Rehabil Psychol 2018 Nov;63(4):563-574. [doi: 10.1037/rep0000241] [Medline: 30247055]

4. Hesdorffer DC. Comorbidity between neurological illness and psychiatric disorders. CNS Spectr 2016 Feb 22;21(3):230-238. [doi: 10.1017/s1092852915000929]
5. Cader Z, Kings J, Kipps C, Langdon D, Mannan R, Marsh S, et al. Transforming community neurology. What commissioners need to know: part A–transformation guide. London, UK: National Health Service; 2016.
6. Cella D, Nowinski C, Peterman A, Victorson D, Miller D, Lai J, et al. The neurology quality-of-life measurement initiative. Arch Phys Med Rehabil 2011 Oct;92(10 Suppl):S28-S36 [FREE Full text] [doi: 10.1016/j.apmr.2011.01.025] [Medline: 21958920]
7. de la Loge C, Dimova S, Mueller K, Phillips G, Durgin TL, Wicks P, et al. PatientsLikeMe® Online Epilepsy Community: Patient characteristics and predictors of poor health-related quality of life. Epilepsy Behav 2016 Oct;63:20-28 [FREE Full text] [doi: 10.1016/j.yebeh.2016.07.035] [Medline: 27544877]
8. Prisnie JC, Sajobi TT, Wang M, Patten SB, Fiest KM, Bulloch AG, et al. Effects of depression and anxiety on quality of life in five common neurological disorders. Gen Hosp Psychiatry 2018 May;52:58-63. [doi: 10.1016/j.genhosppsych.2018.03.009] [Medline: 29684713]
9. Kamenov K, Cabello M, Caballero FF, Cieza A, Sabariego C, Raggi A, et al. Factors Related to Social Support in Neurological and Mental Disorders. PLoS One 2016 Feb 22;11(2):e0149356 [FREE Full text] [doi: 10.1371/journal.pone.0149356] [Medline: 26900847]
10. GBD 2016 Neurology Collaborators. Global, regional, and national burden of neurological disorders, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol 2019 May;18(5):459-480 [FREE Full text] [doi: 10.1016/S1474-4422(18)30499-X] [Medline: 30879893]
11. GBD 2015 Neurological Disorders Collaborator Group. Global, regional, and national burden of neurological disorders during 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet Neurol 2017 Nov;16(11):877-897 [FREE Full text] [doi: 10.1016/S1474-4422(17)30299-5] [Medline: 28931491]
12. Gutierrez-Angel AM, Martinez-Juarez IE, Hernandez-Vanegas LE, Crail-Melendez D. Quality of life and level of burden in primary caregivers of patients with epilepsy: Effect of neuropsychiatric comorbidity. Epilepsy Behav 2018 Apr;81:12-17. [doi: 10.1016/j.yebeh.2018.01.034] [Medline: 29455081]
13. Hlabangana V, Hearn JH. Depression in partner caregivers of people with neurological conditions; associations with self-compassion and quality of life. J Ment Health 2020 Apr 26;29(2):176-181. [doi: 10.1080/09638237.2019.1630724] [Medline: 31241383]
14. Tramonti F, Bonfiglio L, Bongioanni P, Belviso C, Fanciullacci C, Rossi B, et al. Caregiver burden and family functioning in different neurological diseases. Psychol Health Med 2019 Jan 24;24(1):27-34. [doi: 10.1080/13548506.2018.1510131] [Medline: 30141703]
15. Neuro Numbers 2019. Neurological Alliance. 2019 Mar. URL: https://www.neural.org.uk/assets/pdfs/neuro-numbers-2019.pdf [accessed 2022-06-30]
16. Kılınç S, Erdem H, Healey R, Cole J. Finding meaning and purpose: a framework for the self-management of neurological conditions. Disabil Rehabil 2022 Jan 14;44(2):219-230. [doi: 10.1080/09638288.2020.1764115] [Medline: 32406762]
17. World Health Organization (WHO). Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment. Geneva, Switzerland: WHO; 2016.
18. Shegog R, Braverman L, Hixson JD. Digital and technological opportunities in epilepsy: Toward a digital ecosystem for enhanced epilepsy management. Epilepsy Behav 2020 Jan;102:106663. [doi: 10.1016/j.yebeh.2019.106663] [Medline: 31778878]
19. Scholz M, Haase R, Schriefer D, Voigt I, Ziemssen T. Electronic Health Interventions in the Case of Multiple Sclerosis: From Theory to Practice. Brain Sci 2021 Feb 02;11(2):180 [FREE Full text] [doi: 10.3390/brainsci11020180] [Medline: 33540640]
20. Mosadeghi-Nik M, Askari MS, Fatehi F. Mobile health (mHealth) for headache disorders: A review of the evidence base. J Telemed Telecare 2016 Oct 30;22(8):472-477. [doi: 10.1177/1357633x16673275]
21. Klucken J, Krüger R, Schmidt P, Bloem BR. Management of Parkinson's Disease 20 Years from Now: Towards Digital Health Pathways. JPD 2018 Dec 18;8(s1):S85-S94. [doi: 10.3233/jpd-181519]
22. Kettlewell J, das Nair R, Radford K. A systematic review of personal smart technologies used to improve outcomes in adults with acquired brain injuries. Clin Rehabil 2019 Nov 29;33(11):1705-1712. [doi: 10.1177/0269215519865774] [Medline: 31353940]
23. Fischer A, Schröder J, Vettorazzi E, Wolf OT, Pöttgen J, Lau S, et al. An online programme to reduce depression in patients with multiple sclerosis: a randomised controlled trial. The Lancet Psychiatry 2015 Mar;2(3):217-223. [doi: 10.1016/s2215-0366(14)00049-2]
24. Simblett SK, Bruno E, Siddi S, Matcham F, Giuliano L, López JH, RADAR-CNS Consortium. Patient perspectives on the acceptability of mHealth technology for remote measurement and management of epilepsy: A qualitative analysis. Epilepsy Behav 2019 Aug;97:123-129. [doi: 10.1016/j.yebeh.2019.05.035] [Medline: 31247523]
25. Simblett SK, Evans J, Greer B, Curtis H, Matcham F, Radaelli M, RADAR-CNS consortium. Engaging across dimensions of diversity: A cross-national perspective on mHealth tools for managing relapsing remitting and progressive multiple sclerosis. Mult Scler Relat Disord 2019 Jul;32:123-132. [doi: 10.1016/j.msard.2019.04.020] [Medline: 31125754]

26. Escoffery C, McGee R, Bidwell J, Sims C, Thropp EK, Frazier C, et al. A review of mobile apps for epilepsy self-management. Epilepsy Behav 2018 Apr;81:62-69. [doi: 10.1016/j.yebeh.2017.12.010] [Medline: 29494935]
27. Lavorgna L, Brigo F, Moccia M, Leocani L, Lanzillo R, Clerico M, et al. e-Health and multiple sclerosis: An update. Mult Scler 2018 Nov 19;24(13):1657-1664. [doi: 10.1177/1352458518799629] [Medline: 30231004]
28. Marziniak M, Brichetto G, Feys P, Meyding-Lamadé U, Vernon K, Meuth SG. The Use of Digital and Remote Communication Technologies as a Tool for Multiple Sclerosis Management: Narrative Review. JMIR Rehabil Assist Technol 2018 Apr 24;5(1):e5 [FREE Full text] [doi: 10.2196/rehab.7805] [Medline: 29691208]
29. Riggare S, Stamford J, Hägglund M. A Long Way to Go: Patient Perspectives on Digital Health for Parkinson's Disease. J Parkinsons Dis 2021;11(s1):S5-S10 [FREE Full text] [doi: 10.3233/JPD-202408] [Medline: 33682728]
30. Stubberud A, Linde M. Digital Technology and Mobile Health in Behavioral Migraine Therapy: a Narrative Review. Curr Pain Headache Rep 2018 Jul 31;22(10):66. [doi: 10.1007/s11916-018-0718-0] [Medline: 30066141]
31. Audulv, Packer T, Hutchinson S, Roger KS, Kephart G. Coping, adapting or self-managing - what is the difference? A concept review based on the neurological literature. J Adv Nurs 2016 Nov 27;72(11):2629-2643. [doi: 10.1111/jan.13037] [Medline: 27272388]
32. Babbage DR, van Kessel K, Drown J, Thomas S, Sezier A, Thomas P, et al. MS Energize: Field trial of an app for self-management of fatigue for people with multiple sclerosis. Internet Interv 2019 Dec;18:100291 [FREE Full text] [doi: 10.1016/j.invent.2019.100291] [Medline: 31890637]
33. Bromberg J, Wood ME, Black RA, Surette DA, Zacharoff KL, Chiauzzi EJ. A randomized trial of a web-based intervention to improve migraine self-management and coping. Headache 2012 Mar;52(2):244-261 [FREE Full text] [doi: 10.1111/j.1526-4610.2011.02031.x] [Medline: 22413151]
34. DasMahapatra P, Chiauzzi E, Pujol LM, Los C, Trudeau KJ. Mediators and moderators of chronic pain outcomes in an online self-management program. Clin J Pain 2015 May;31(5):404-413 [FREE Full text] [doi: 10.1097/AJP.0000000000000125] [Medline: 24918473]
35. Ghahari S, Packer T. Effectiveness of online and face-to-face fatigue self-management programmes for adults with neurological conditions. Disabil Rehabil 2012 Oct 12;34(7):564-573. [doi: 10.3109/09638288.2011.613518] [Medline: 21988717]
36. Sorbi M, Kleiboer A, van Silfhout H, Vink G, Passchier J. Medium-term effectiveness of online behavioral training in migraine self-management: A randomized trial controlled over 10 months. Cephalalgia 2015 Jun 16;35(7):608-618. [doi: 10.1177/0333102414547137] [Medline: 25228685]
37. Sorbi MJ, van der Vaart R. User acceptance of an Internet training aid for migraine self-management. J Telemed Telecare 2010 Jan 19;16(1):20-24. [doi: 10.1258/jtt.2009.001007] [Medline: 20086263]
38. van Kessel K, Wouldes T, Moss-Morris R. A New Zealand pilot randomized controlled trial of a web-based interactive self-management programme (MSInvigor8) with and without email support for the treatment of multiple sclerosis fatigue. Clin Rehabil 2016 May 07;30(5):454-462. [doi: 10.1177/0269215515584800] [Medline: 25952587]
39. DiIorio C, Escoffery C, McCarty F, Yeager KA, Henry TR, Koganti A, et al. Evaluation of WebEase: an epilepsy self-management Web site. Health Educ Res 2009 Apr 21;24(2):185-197. [doi: 10.1093/her/cyn012] [Medline: 18682382]
40. Nielsen AS, Halamka JD, Kinkel RP. Internet portal use in an academic multiple sclerosis center. J Am Med Inform Assoc 2012 Jan 01;19(1):128-133 [FREE Full text] [doi: 10.1136/amiajnl-2011-000177] [Medline: 21571744]
41. Pierce LL, Steiner V. Usage and design evaluation by family caregivers of a stroke intervention web site. J Neurosci Nurs 2013 Oct;45(5):254-261 [FREE Full text] [doi: 10.1097/JNN.0b013e31829dba61] [Medline: 24025464]
42. Demir Y, Gozum S. Evaluation of Quality, Content, and Use of the Web Site Prepared for Family Members Giving Care to Stroke Patients. Comput Inform Nurs 2015 Sep;33(9):396-403. [doi: 10.1097/CIN.0000000000000165] [Medline: 26125455]
43. Boeschoten RE, Nieuwenhuis MM, van Oppen P, Uitdehaag BM, Polman CH, Collette EH, et al. Feasibility and outcome of a web-based self-help intervention for depressive symptoms in patients with multiple sclerosis: a pilot study. J Neurol Sci 2012 Apr 15;315(1-2):104-109 [FREE Full text] [doi: 10.1016/j.jns.2011.11.016] [Medline: 22133479]
44. Kleiboer A, Sorbi M, van Silfhout M, Kooistra L, Passchier J. Short-term effectiveness of an online behavioral training in migraine self-management: a randomized controlled trial. Behav Res Ther 2014 Oct;61:61-69. [doi: 10.1016/j.brat.2014.07.009] [Medline: 25133856]
45. Begley C, Shegog R, Harding A, Goldsmith C, Hope O, Newmark M. Longitudinal feasibility of MINDSET: a clinic decision aid for epilepsy self-management. Epilepsy Behav 2015 Mar;44:143-150. [doi: 10.1016/j.yebeh.2014.12.031] [Medline: 25705825]
46. DiIorio C, Bamps Y, Walker ER, Escoffery C. Results of a research study evaluating WebEase, an online epilepsy self-management program. Epilepsy Behav 2011 Nov;22(3):469-474. [doi: 10.1016/j.yebeh.2011.07.030] [Medline: 21889413]
47. Walker ER, Bamps Y, Burdett A, Rothkopf J, Diiorio C. Social support for self-management behaviors among people with epilepsy: a content analysis of the WebEase program. Epilepsy Behav 2012 Mar;23(3):285-290. [doi: 10.1016/j.yebeh.2012.01.006] [Medline: 22364762]

48. D'hooghe M, Van Gassen G, Kos D, Bouquiaux O, Cambron M, Decoo D, et al. Improving fatigue in multiple sclerosis by smartphone-supported energy management: The MS TeleCoach feasibility study. Mult Scler Relat Disord 2018 May;22:90-96 [FREE Full text] [doi: 10.1016/j.msard.2018.03.020] [Medline: 29649789]
49. Miller DM, Moore SM, Fox RJ, Atreja A, Fu AZ, Lee J, et al. Web-based self-management for patients with multiple sclerosis: a practical, randomized trial. Telemed J E Health 2011 Jan;17(1):5-13 [FREE Full text] [doi: 10.1089/tmj.2010.0133] [Medline: 21214498]
50. Schneider MA, Howard KA. Using Technology to Enhance Discharge Teaching and Improve Coping for Patients After Stroke. J Neurosci Nurs 2017 Jun;49(3):152-156. [doi: 10.1097/JNN.0000000000000275] [Medline: 28346312]
51. Greiner P, Sawka A, Imison E. Patient and Physician Perspectives on MSdialog, an Electronic PRO Diary in Multiple Sclerosis. Patient 2015 Dec 8;8(6):541-550 [FREE Full text] [doi: 10.1007/s40271-015-0140-1] [Medline: 26350792]
52. Hixson JD, Barnes D, Parko K, Durgin T, Van Bebber S, Graham A, et al. Patients optimizing epilepsy management via an online community. Neurology 2015 Jun 17;85(2):129-136. [doi: 10.1212/wnl.0000000000001728]
53. Beck AT, Steer RA, Brown GK. Beck Depression Inventory II. Psychological Assessment. The Psychological Corporation: San Antonio, TX; 1996.
54. Lovibond S, Lovibond P. Manual for the Depression Anxiety Stress Scales. 2nd Edition. Sydney, NSW, Australia: Psychology Foundation; 1995.
55. 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]
56. Radloff LS. The CES-D Scale. Applied Psychological Measurement 2016 Jul 26;1(3):385-401. [doi: 10.1177/014662167700100306]
57. Ghahari S, Leigh Packer T, Passmore AE. Effectiveness of an online fatigue self-management programme for people with chronic neurological conditions: a randomized controlled trial. Clin Rehabil 2010 Aug 11;24(8):727-744. [doi: 10.1177/0269215509360648] [Medline: 20543022]
58. Pierce LL, Steiner VL, Khuder SA, Govoni AL, Horn LJ. The effect of a Web-based stroke intervention on carers' well-being and survivors' use of healthcare services. Disabil Rehabil 2009 Sep 15;31(20):1676-1684. [doi: 10.1080/09638280902751972] [Medline: 19479528]
59. Penner I, Raselli C, Stöcklin M, Opwis K, Kappos L, Calabrese P. The Fatigue Scale for Motor and Cognitive Functions (FSMC): validation of a new instrument to assess multiple sclerosis-related fatigue. Mult Scler 2009 Dec 07;15(12):1509-1517. [doi: 10.1177/1352458509348519] [Medline: 19995840]
60. Fisk JD, Pontefract A, Ritvo PG, Archibald CJ, Murray T. The Impact of Fatigue on Patients with Multiple Sclerosis. Can. j. neurol. sci 2015 Sep 18;21(1):9-14. [doi: 10.1017/s0317167100048691]

Abbreviations

CASP: Critical Appraisal Skills Programme
mHealth: mobile health
MS: multiple sclerosis
NHS: National Health Service

Edited by A Mavragani; submitted 11.07.21; peer-reviewed by K Sato, A Saad, C Carroll; comments to author 19.10.21; revised version received 13.03.22; accepted 19.05.22; published 28.07.22.

Please cite as:
Spreadbury JH, Young A, Kipps CM
A Comprehensive Literature Search of Digital Health Technology Use in Neurological Conditions: Review of Digital Tools to Promote Self-management and Support
J Med Internet Res 2022;24(7):e31929
URL: https://www.jmir.org/2022/7/e31929
doi:10.2196/31929
PMID:35900822

Viewpoint

StudyU: A Platform for Designing and Conducting Innovative Digital N-of-1 Trials

Stefan Konigorski[1,2,3], PhD; Sarah Wernicke[1,2,3], MSc; Tamara Slosarek[1,2,3], MSc; Alexander M Zenner[1,2], MSc; Nils Strelow[1,2], MSc; Darius F Ruether[1,2,4], Dr med; Florian Henschel[1,2], MSc; Manisha Manaswini[1,2], MSc; Fabian Pottbäcker[1,2], BSc; Jonathan A Edelman[1,5], PhD; Babajide Owoyele[1], MSc; Matteo Danieletto[3], PhD; Eddye Golden[3], MPH; Micol Zweig[3], MPH; Girish N Nadkarni[3], MD, MPH; Erwin Böttinger[1,2,3], MD

[1]Digital Health Center, Hasso Plattner Institute for Digital Engineering, University of Potsdam, Potsdam, Germany

[2]Digital Engineering Faculty, University of Potsdam, Potsdam, Germany

[3]Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[4]Department of Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

[5]The Center for Advanced Design Studies, Palo Alto, CA, United States

Corresponding Author:
Stefan Konigorski, PhD
Digital Health Center
Hasso Plattner Institute for Digital Engineering
University of Potsdam
Prof.-Dr.-Helmert-Straße 2-3
Potsdam, 14482
Germany
Phone: 49 331 5509 4873
Email: stefan.konigorski@hpi.de

Abstract

N-of-1 trials are the gold standard study design to evaluate individual treatment effects and derive personalized treatment strategies. Digital tools have the potential to initiate a new era of N-of-1 trials in terms of scale and scope, but fully functional platforms are not yet available. Here, we present the open source StudyU platform, which includes the *StudyU Designer* and StudyU app. With the *StudyU Designer*, scientists are given a collaborative web application to digitally specify, publish, and conduct N-of-1 trials. The StudyU app is a smartphone app with innovative user-centric elements for participants to partake in trials published through the *StudyU Designer* to assess the effects of different interventions on their health. Thereby, the StudyU platform allows clinicians and researchers worldwide to easily design and conduct digital N-of-1 trials in a safe manner. We envision that StudyU can change the landscape of personalized treatments both for patients and healthy individuals, democratize and personalize evidence generation for self-optimization and medicine, and can be integrated in clinical practice.

(J Med Internet Res 2022;24(7):e35884) doi:10.2196/35884

KEYWORDS

digital interventions; N-of-1 trial; SCED; single-case experimental design; web application; mobile application; app; digital health

Introduction

A widespread aim in current medical research is to derive personalized treatment strategies, which are at the heart of treating every single patient with the best possible therapies. One motivation underlying this aim is that many drugs are only effective in up to 50% of patients [1-3]. Hence, treatment guidelines based on population-level randomized controlled trials (RCTs), which derive the best average treatment, may result in ineffective treatment or side effects in up to 50% of patients. As another characteristic of traditional RCTs, study participants only provide data for population-level analyses and do not profit from participation in the studies. Population-level RCTs are not meant to provide insights for individual participants. The gold standard design for evaluating individual-level treatment effects is prospective longitudinal N-of-1 trials [4], which are multi-crossover RCTs of sample size one [5,6]. That is, the study participant is administered the treatments of interest over time in accordance with a predefined setup of treatment length, duration, treatment blocks, and washout phases. In the literature, N-of-1 trials are sometimes

used as a synonym for single-case experimental designs (SCEDs), mainly in the United Kingdom, but mostly used to describe a special case of SCED [7].

N-of-1 trials are well suited for studies when there is large interindividual heterogeneity in treatment effects, as well as when there are subpopulation groups or individuals with comorbidities of interest, who have been excluded from population-level RCTs. Important considerations relate to blinding, randomization, the typically arising correlation of measurements over time, and carryover effects of interventions, which can be approached either through the design of N-of-1 trials or in the subsequent statistical analysis. Series of N-of-1 trials can be aggregated to provide population-level estimates of treatment effects with similar efficiency to that of RCTs but require smaller numbers of participants [4,8,9]. Historically, there have been local implementations of N-of-1 trials in hospitals in the United States, Canada, and Australia [4,10]; series of articles on N-of-1 trials have been published in medical and epidemiological journals [11,12]; and networks on N-of-1 studies have been formed [13]. The advancements in digital technologies provide the potential to initiate a new era of N-of-1 trials in terms of scale and scope and have opened up new avenues to offer remote health care. In particular, performing N-of-1 trials digitally allows a seamless integration of trials in daily life, which can save time for participants and researchers since there is no need to visit a study center. This is especially important if daily measurements are collected. Extensive data can also be assessed passively to further reduce the burden on the participant if sensors are linked to a digital N-of-1 trial app. Finally, recruitment of participants can be simplified as participants from all over the world can participate in a trial, which is even more important for rare diseases where there are few potential participants [14].

Nonetheless, N-of-1 trials have not been integrated into mainstream clinical research or clinical practice. One of the underlying reasons might be that despite recently published guidelines [15-17], there are still considerations of the ethical framework when applying N-of-1 trials in clinical care [18]. Typically, series of N-of-1 trials designed with a specific research aim require ethics approval, while single N-of-1 trials with a clinical aim for a single patient do not require it. Often, however, this distinction is not clear. For example, let us consider a potential N-of-1 trial where physicians have the goal of finding out whether a particular drug is effective in off-label use for patients with chronic conditions such as chronic liver disease. The study setting involves patients treated in a specialized clinic department. In addition to treating the patients, the physicians in charge might be interested in knowing if the results are generalizable. In this situation, a series of N-of-1 trials can be designed, comparing standard care to standard care plus off-label drug use over different crossover periods. This example shows how N-of-1 trials can be woven into clinical practice and, as such, how their innovative design might be of high interest to physicians, if they can be performed easily. The latter point presents maybe the most important reason why N-of-1 trials have not been picked up more broadly since there is no platform available that allows for an easy and large-scale implementation of digital N-of-1 trials. As of now, conducting a digital N-of-1 trial generally necessitates the development of a new app.

Here, we present *StudyU*. *StudyU* provides an open-source, free, and easy-to-use platform with a study designer app for researchers, which allows easy design, customization, and implementation of N-of-1 trials, as well as a study participant app that allows participants to partake in these trials without having to set up user accounts. This serves to allow novel interactions among researchers, trials, and study participants.

Design and Methods

Related Work

A number of apps for conducting N-of-1 trials have been published. In order to attract physicians and researchers to design and conduct digital N-of-1 trials through a platform, the platform has to be accessible as easily as possible, should be able to implement interventions beyond mere symptom tracking, and, more generally, should allow designing studies flexibly for different interventions and different outcomes. Finally, analyzing the results in the app and providing the results back to the participant is an essential component to use the intrinsic patient-empowering potential of N-of-1 trials.

Table 1 presents, to the best of our knowledge, an overview of the most relevant apps that can be used to perform individual-level studies, particularly for N-of-1 trials. We want to note that it does not provide an exhaustive overview, and some other commercial platforms exist with limited publicly available information on their functionality; for example, the N of 1 platform by Digital Infuzion [19], which allows observational tracking of study participants. Furthermore, some apps have been developed, which focus on the cocreation of single N-of-1 trials by study participants themselves [20,21], which are not the focus here.

The Trialist app [22] provides results back to the participants and has been used in different N-of-1 trials but is currently not available for download and general use. The N1 app [23] had one study implemented for iOS users in the United States, which investigated the effects of caffeine and L-theanine on cognitive outcomes. It did not allow customization and further implementation of studies and is currently not available. Several apps provide functionalities for self-tracking and self-quantification but do not allow for an experimental evaluation of interventions (eg, mPower [24] and Parkinson mPower2 [24]). Of these apps, N1 and mPower are based on the Apple Research Kit. OpenClinica [25] allows creating and conducting studies but focuses on electronic data capture and data management and neither reports results to participants nor allows a collaborative creation of studies. QuantifyMe [26] is a platform that allows users to choose from a limited and prespecified set of interventions and outcomes to design a study without further customization possibilities. TummyTrials [27] and SleepCoacher [28] provide possibilities to choose from a set of specified interventions and investigate their effect on sleep and on food triggers in irritable bowel syndrome, respectively. Finally, PACO [29] and movisensXS [30,31] provide tools to design studies but are missing the main

component of N-of-1 trials, in that the study app only gathers data—the results are not analyzed in the app and are not reported back to the study participant in the app. PACO has a further restriction that it is only available outside of the European Union and Switzerland. Furthermore, all of these mentioned platforms, except TummyTrials, require user accounts, which can create difficulties in terms of data privacy, especially if apps are planned to be used in different countries.

Table 1. Overview of existing apps and platforms that are suitable for gathering individual-level data. Some report the results of the conducted studies back to the user (column "Statistical evaluation of results").

Name	App availability	Possible studies/ diseases	Platforms	Statistical evaluation of results	Customizable	Able to perform N-of-1 trials	Requires a user account	Link to the software
Trialist	No	Multiple options (for chronic pain only)	iOS, Android, or web	Yes	Limited options	Yes	Yes	N/A[a]
mPower	Only United States	1 (linked to Parkinson)	iOS	No	No	No	Yes	[32]
Parkinson mPower2	Yes	1 (linked to Parkinson)	iOS	No	No	No	Yes	[33]
PACO	Outside of the European Union and Switzerland	Flexible creation	iOS, Android, and web	No	Yes	Yes	Yes	[34]
movisensXS	Yes	Flexible	Android	No	Yes	Yes	Yes	[35]
OpenClinica	Yes	0	Web	No	Yes	Yes	Yes	[36]
N1	Only United States	1 (linked to cognitive health)	iOS	Yes	No	Yes	Yes	N/A
QuantifyMe	Source code only	4	Android	Yes	Limited options	Yes	Yes	[37]
TummyTrials	Source code only	4 (linked to irritable bowel syndrome)	iOS	Yes	Limited options	Yes	No	[38]
SleepCoacher	Yes	Multiple options (linked to sleep)	iOS and Android	Yes	Limited options	Yes	Yes	[39]
StudyU	Yes	Flexible creation	iOS, Android, and web	Yes	Yes	Yes	No	[40-42]

[a]N/A: not applicable.

Vision

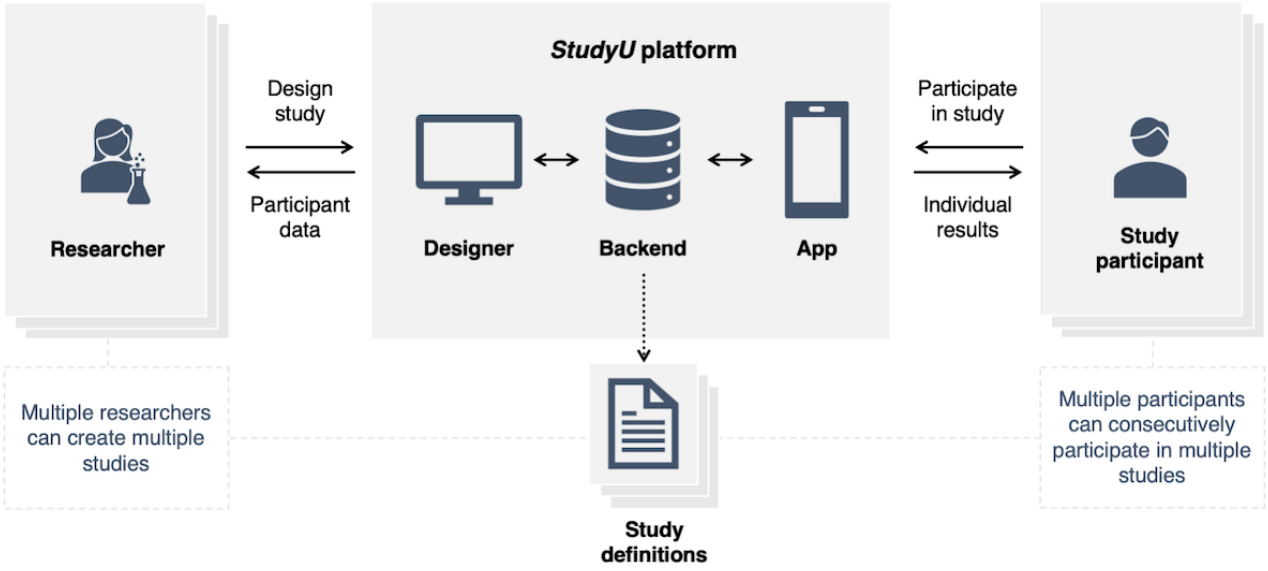
With *StudyU*, our goal is to attract more study participants and researchers to conduct and participate in N-of-1 trials by reducing the set-up process and implementation efforts. We envision that health scientists, medical researchers, and physicians worldwide can use it to collaboratively design and conduct N-of-1 trials. *StudyU* can therefore serve as a platform to contribute to open, transparent, and reproducible medical science by (1) making the study designs of different designed trials directly available to foster reproducibility and well-designed studies, and (2) making the anonymized data contributed by the study participants of the platform available for analysis to foster the generation of novel medical insights on health intervention effects at the individual and population levels. We envision enabling democratization and personalization for evidence generation in medicine and personal self-optimization.

The StudyU Platform

The *StudyU* platform consists of 3 main parts, as illustrated in Figure 1 (see Supplementary Text 1 in Multimedia Appendix 1 for more details on the architecture):

1. the *StudyU Designer* web application for researchers,
2. the *StudyU* app for mobile devices, and
3. the backend where the participant data, study definitions, etc, are safely stored.

Figure 1. Architecture of the StudyU platform. Multiple researchers can collaboratively design and create studies and publish them. Then, study participants can partake in published studies. Study definitions and participant study data are stored in the backend.



Designing and implementing a study with the *StudyU Designer* includes specifying the interventions (Multimedia Appendix 1, p19), eligibility criteria (Multimedia Appendix 1, p20), observations (Multimedia Appendix 1, p21), how they are scheduled, computing the results and displaying them to the participant (see *App*), and consent (Multimedia Appendix 1, p22). Such designed studies are then available to participants through the *StudyU* app. This user journey is described in more detail in Multimedia Appendix 1, Supplementary Text 2. The designer and the app are currently available in German and English, with apps in Spanish, French, and Korean planned in the near future. In the following sections, the technical setup and the main parts of *StudyU* are described.

Technical Setup and Use

The *StudyU* frontend applications are written in Flutter [43], an open-source, cross-platform user interface framework by Google based on the Dart programming language. With this, one single code base can be compiled to performant applications for multiple platforms: mobile, web, and desktop. Parse [44] is used as a backend, which is a platform that incorporates various functionalities such as object storage, user authentication, and push notifications. All components are organized and composed as Docker [45] containers for easy deployment. The source code for the *StudyU* platform is publicly and freely available on GitHub [40], and the *StudyU* app is available on Google Play and the Apple App Store [41,42]. For demonstration purposes, the backend is deployed on Back4App [46], and the frontend applications of the *StudyU Designer* and *StudyU* app are deployed on Google Cloud Run [47,48]. *StudyU* can also be deployed into any HIPAA (Health Insurance Portability and Accountability Act)–compliant and GDPR (General Data Protection Regulation)–compliant cloud system.

In the current implementation of *StudyU*, two choices can be made regarding how to use the platform, which provides flexibility to the needs of the researcher. First, *StudyU* can either be installed on one's own separate server (or cloud) instance, or *StudyU* can be run and accessed on a central server operated by a third party. Second, studies can be designed and published individually or in collaboration with other researchers from other institutions. For collaborative design, studies can be accessed, edited, and saved by multiple researchers from multiple institutions. The studies can be accessed by multiple researchers at the same time, with the restriction that only one researcher can save data at the same time.

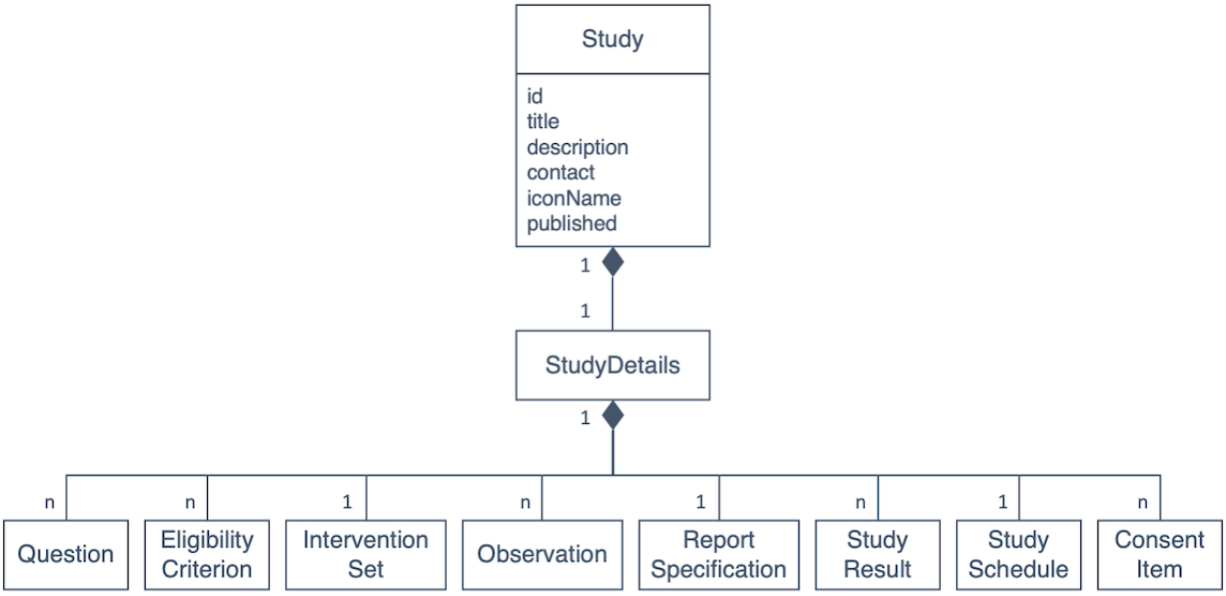
Study Model

StudyU is based on a generic study representation, which is essential to dynamically support multiple studies. The representation encompasses study metadata and study details. The metadata of a study include basic information such as the title, a short description, and the researcher's contact, including the name of the institutional review board (IRB) and protocol number. The study details contain all information that is needed to execute the study: eligibility questions and criteria, interventions, observations, specification of output and report data, schedule, and consent. All objects and relationships are serialized and stored in JavaScript Object Notation (JSON) format. The overall components of this study model are displayed in Figure 2 (see Multimedia Appendix 1, p23 and Supplementary Text 3 for more details).

The generic study model allows the design of many different N-of-1 trials in *StudyU*. This is illustrated in 2 example studies that are implemented in *StudyU*:

1. Investigation and comparison of the effect of any 2 of the following daily interventions on the intensity of chronic low back pain: willow bark tea, arnica balm, and warming pad
2. Investigation of the effect of any 2 of the following daily interventions on diffuse abdominal pain in irritable bowel syndrome: gluten-free diet, low-fiber diet, and fructose-free diet.

Figure 2. A simplified overview of the StudyU study model. The notation is based on the Unified Modeling Language class diagram notation, which defines properties of single classes in rectangles and associations between multiple classes as connections. The associations shown in this diagram with a filled diamond at one end mean that one class, for example, "Study," is composed of another class, in this case, "StudyDetails." Numbers shown at associations indicate how many instances of one class take part in this association, for example, n "Observation" objects can be associated with one "StudyDetails" object.

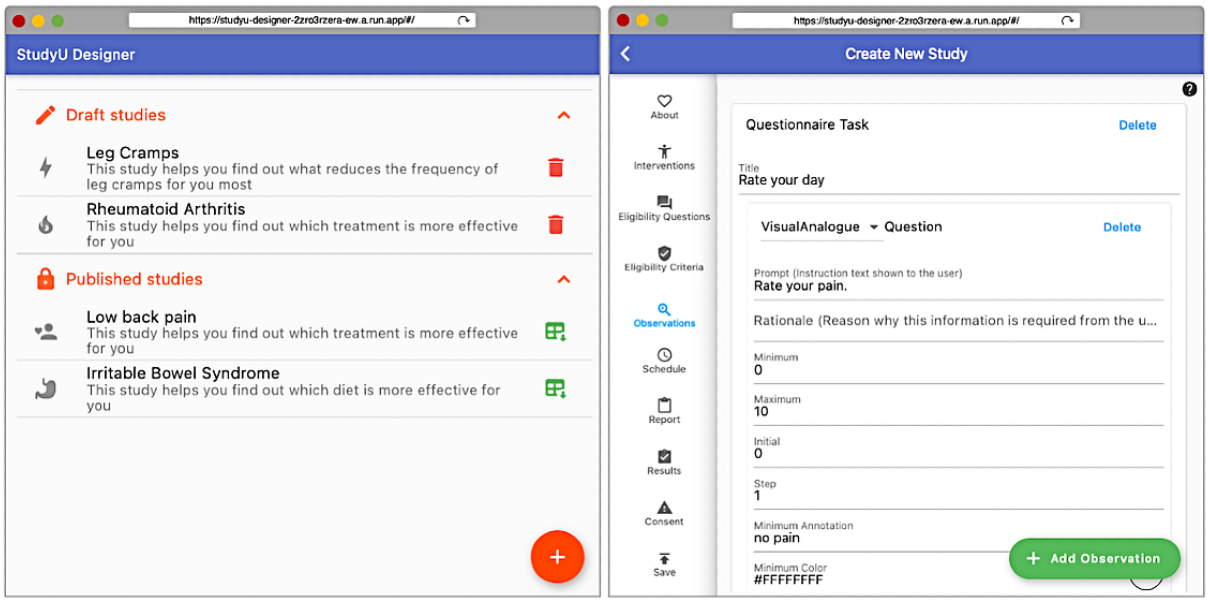


StudyU Designer

The *StudyU Designer* consists of 2 main components: the dashboard and the editor. The rationale behind this concept is to build a user-friendly tool for researchers, which provides a logical framework with all the necessary components to plan and conduct a study. Figure 3 shows the dashboard, which displays drafted studies and published studies. Once a study is published, it is available to users in the app and cannot be edited anymore in the designer. For published studies, researchers can download participant data in comma-separated values (CSV) format.

When adding a new study or editing a draft study, the editor leads through all study specifications as defined in the study model, such as interventions, observations, inclusion and exclusion criteria, consent, the format of the downloadable CSV file with study results, and the specification of reports shown to the user in the app. More editor examples and more details are shown in Multimedia Appendix 1, Supplementary Text 4. The sole responsibility for studies lies with the study designers, and in order to ensure the study participants' appropriateness and safety of studies published in *StudyU*, the terms of use of *StudyU* prohibit misuse of the platform and require that researchers have conducted training on good clinical practice. Researchers have to include an IRB protocol number in the study metadata to assure participants of the adequacy of their study.

Figure 3. The dashboard of the *StudyU Designer* with drafted and published studies and an editor screen for observation definitions.



App

The app enables users to participate in all studies that were created and published in the designer. This has the major advantage that participants do not have to download multiple apps for different studies but can partake in different studies through the same app and have an overview of all of them. After the welcome screen (Figure 4A), users can select which of the published studies they want to participate in. Before enrolling in one study, the study metadata are displayed in the study overview screen (Figure 4B). Then, users are led through the onboarding process with a validation of their eligibility, intervention selection, and declaration of consent. Finally, users arrive at the overview of daily tasks (Figure 4C), which contains the study progress bar and the daily tasks (Figure 4D). This is also the default screen users see when opening the app after the initial onboarding.

A centerpiece of the *StudyU* app is the result visualization, which is illustrated in Figure 5. In order to ensure that no biases occur after having viewed the results, participants can only view them upon completion of a minimum study length specified by the researcher. For this purpose, a recommended study length is displayed to researchers in the designer, which should be calculated on the basis of a statistical sample size calculation. It should be noted that in the current demonstration of *StudyU* [48], the results are available from the first day in order to visualize the results. Through progress bars, the current status of the participant in the study is visualized to show how many more observations are needed; the effects of the interventions (if present) can be expected with the specified statistical power if the participants continue with the intervention at least until they reach the minimum study duration and report the measurements without missing data. In the study designer, different report types can be selected: (1) the visualization of a linear regression model that tests if the intervention has an effect on the outcome or (2) the report and explanation of individual results to the participant in bar charts. The definition of report types is implemented in an extensible way. More details are provided in Multimedia Appendix 1, Supplementary Text 5.

Figure 4. Initial screens of the StudyU app including the welcome screen, study overview screen, and daily screens.

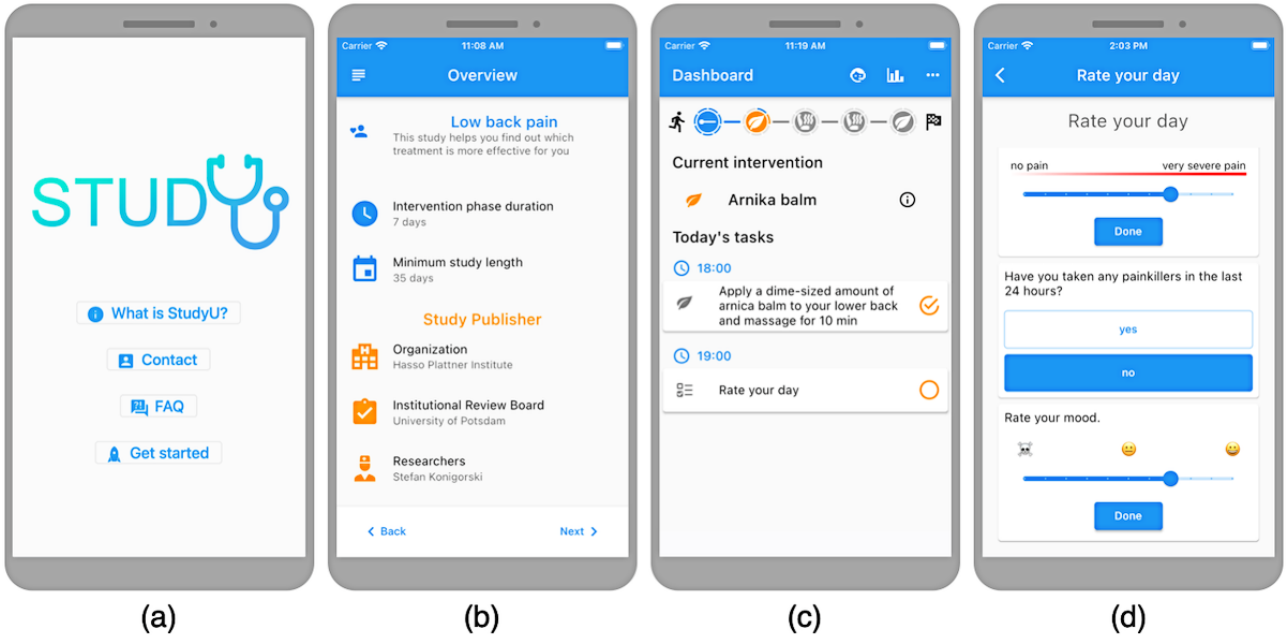
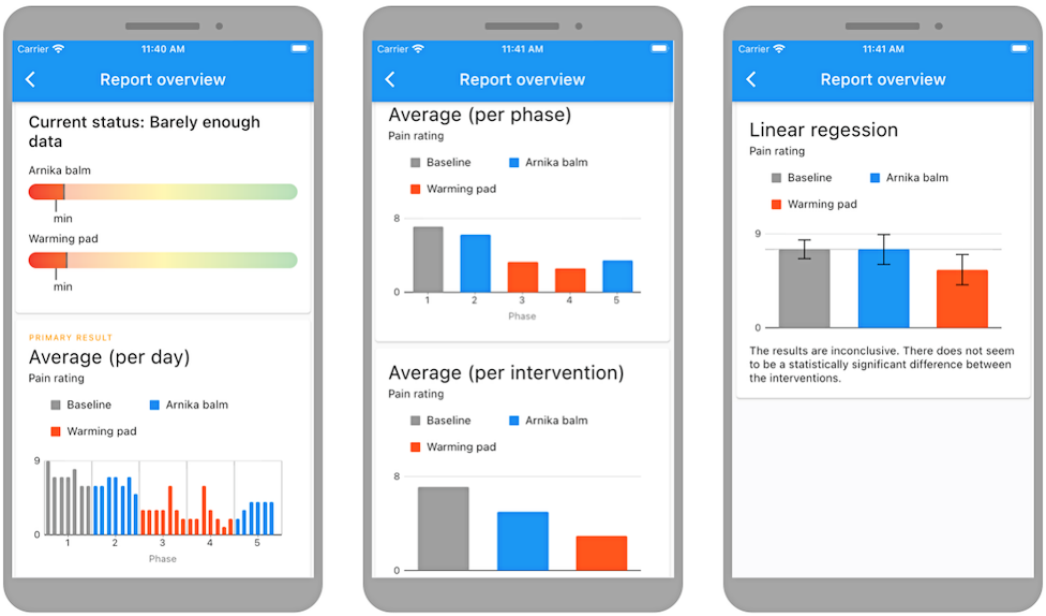


Figure 5. Examples of study reports. The power bar in the top-left panel indicates whether enough data were collected to observe an effect. Reports are displayed either as a linear regression report or as a report showing data aggregated by day, phase, or intervention.



Data Processing

The studies carried out on the *StudyU* platform adhere to applicable ethical principles and international regulations, in particular, the GDPR (European Union) 2016/679 [49]. When the participant opens the app and accepts the terms of service, a new anonymous user account is created with a random ID that is assigned to the participant. Thus, no user profile is needed—which could be used to identify the participant—there is no log-in requirement for the app, and the participant does not need to create a password. The anonymous account is saved in the backend and on the device. Whenever the app is opened, the anonymous account also gets activated. If the participant completes a daily task, the results are stored inside the user study object and updated on the server. With this setup, there is no risk of data loss due to the study participant logging out of the app or forgetting a password. The only risk is losing the smartphone, in which case the anonymous link cannot be recovered. There is the possibility to opt out of the study, which deletes the unfinished study and the local storage and reference to it. The participant can also choose for his/her data to be deleted locally and on the server. These are some important principles of good clinical practice, and we further require every researcher using the *StudyU Designer* explicitly to have followed training in good clinical practice.

The legal basis for processing the study data is the consent provided by the participant via the researcher-defined consent form. Researchers can link and analyze data in different ways in the backend using random user IDs but cannot link them to specific participants. The setup of *StudyU* does not collect identifiable information, and we also discourage researchers from assessing information which includes participant-identifiable data in their designed studies. We anticipate that this setup without user accounts will satisfy the regulations and data security standards in most countries, allowing a broad use of *StudyU*.

Discussion

Here, we have presented the *StudyU* platform, which allows researchers to undertake an easy design, customization, and implementation of N-of-1 trials and allows individuals to participate in those trials without having to set up user accounts. Through the *StudyU Designer*, researchers can collaboratively design trials. *StudyU* is available open-source for iOS, Android, and the web, is free to use, and provides anonymized data entry, which prevents tracing back the data to the participants. It allows conducting the entire study process digitally: study design, participant recruitment, inclusion and exclusion of participants through the study app, automatically analyzing the individual data in the app and reporting the results back to the participant, and saving the data in the secure backend so that researchers can analyze it further and aggregate it across N-of-1 trials. As further innovative concepts, we provide electronic consent and the possibility for the study participants to view their progress through the study on a progress bar. With these features, *StudyU* is currently the only available platform that allows flexibility to the N-of-1 trial design and the capacity to conduct them completely digitally. All other existing apps have limitations in the platforms they support, the possibility and customizability to design individual trials, the freedom to use the app without having to set up a user account, and the automated in-app statistical analysis to provide the results back to the participant.

As participants likely start the trial with high intrinsic motivation and high expectations of getting insights into their health, it is critical that they are not disappointed and drop out of the trial. The progress bar keeps the participant informed when they have reached the targeted study length and when they can view a statistical evaluation of their results. We expect that it can also encourage them to continue the study for a longer time before viewing the results in order to estimate more precise treatment effects, thereby extending the classical statistical power–based sample size calculation. With this, we envision that participants understand the value of long-term participation in the study and

stay motivated for a longer time so that dropout rates can be decreased, thus mixing elements of extrinsic motivation to intrinsic motivation [50]. Regarding the use of progress bars, there is some conflicting evidence in the literature [51,52] regarding whether they have a positive effect on increasing adherence, with some suggestions that only specific types of progress bars (ie, fast-to-slow presentation) are beneficial. For this reason, we propose a new approach of the progress bar in future iterations of *StudyU* to offer participants the chance to look at the results at any time if they want, with the caveat that their results will only be statistically evaluated once to avoid biased results. We expect positive effects on study adherence of such a design, similar to the endowed progress effect [53].

Using the *StudyU* platform, N-of-1 trials can be designed not only to study the effect of many different health interventions and lifestyle factors on health outcomes in rare diseases and chronic diseases, but also to evaluate the effectiveness of digital health apps. N-of-1 trials can evaluate the effect of health interventions truly in the real-world setting. Especially with the ongoing COVID-19 pandemic highlighting the importance of remote and digital medicine, evaluation and digital integration into the home environment are of high value. While fully digital trials in the home environment can provide challenges for N-of-1 trials owing to possible carryover or confounding effects that have to be considered in the automated analysis, the challenges can be addressed through the implementation of more advanced statistical and machine learning methods. In fact, recent years have seen an unprecedented development of deep learning methods for estimating the individual-level effects of health interventions from population-level studies and to predict individual disease trajectories and individual treatment effects [54-58]. These methods are often based on nontestable assumptions, require large data sets, and have limitations in their interpretability of individual treatment effects in complex causal graphs. Combining them with the design advantages of N-of-1 trials can help derive fully automated analyses of complex real-life trials.

Two important considerations in N-of-1 trials are randomization and blinding to ensure that unbiased estimates of causal effects can be obtained. Randomizing study designs in a within-persons (eg, the order of treatment A and B within each cycle) and a between-persons manner can be implemented in *StudyU* if desired, but it should be considered that a deterministic sequence might be able to counterbalance specific time-confounding effects for a given participant, while a randomized sequence can achieve this on average. Blinding can occur on 2 levels: blinding researchers to treatment allocation and blinding study participants to treatment allocation. Researchers can be blinded in *StudyU* by incorporating another person who controls the allocation in the design of the trial. Blinding of study participants with respect to which interventions they are currently following is not possible in many digital N-of-1 trials, as, for example, drinking tea or using a warming patch are visibly different. Blinding would have to be achieved with help of a researcher, physician, or third person and can be implemented in *StudyU* by naming interventions anonymized A and B and providing, for example, similar looking pills for A and B. Such blinding can prevent biases that might arise from the participants during the trial. However, this has to be balanced with the aim of N-of-1 trials to benefit and empower the participant. More importantly, it should be remembered that any conclusion that intervention A works better than B for a given participant also holds for any nonblinded trial—we only do not know the extent to which this effect was due to the intervention or accompanying beliefs. However, the participant might not care why the intervention worked but might rather care about the fact that it worked.

We plan to include several extensions in *StudyU* in the future. First, for the study designer, we plan to add more features encouraging the collaboration on study designs. Setting up a database of interested researchers, clinicians, and institutions can help search for partners in designing and conducting the studies. In the current version of *StudyU*, all studies are, by default, public to enhance full collaboration and allow for open-access study development. We are working on a more fine-grained collaboration platform, which allows the researcher to make both the creation and conducting of studies fully public or private to a selected group of collaborators and selected group of invited study participants. As a second new feature in *StudyU*, we will include the possibility to link sensor-based data to measure health outcomes and covariates and also allow the integration of other digital health apps in *StudyU*. Third, we will provide the possibility to design adaptive trials, for example, including elements from microrandomized trials and just-in-time adaptive interventions [59-61]. Fourth, we plan to implement a more elaborate progress bar visualizing the study progress of the participants. The current progress bar is based on the study duration and number of past measurements, but a more exact measurement would be to focus on the number of nonmissing measurements. This feature can be added by including an automated check for the validity and completeness of the recorded data and feeding it back into the progress bar. Fifth, we plan to integrate more complex statistical and machine learning methods in the study app so that complex individual-level treatment effects of potentially time-varying treatments and time-varying confounders can be included in the modeling and result reporting to the individuals. Currently, only linear regression and *t* tests are implemented in *StudyU*. They provide simple models with easily interpretable results and have been shown to provide efficient and robust treatment effect estimates even when autocorrelation and time trends are present [62]. Nonetheless, implementing more complex statistical models such as Bayesian mixed models or G-estimation will allow a more fine-grained and powerful analysis. Finally, we are working on the development of user-centric N-of-1 trials designed by the study participants themselves and are excited to integrate these study designs as well as the study results into *StudyU* [20]. We envision that placing a higher focus on the cocreation of trials with participants can be very important to increase adherence to the trial, especially for long-term experiments, which is not straightforward as shown in other studies [21,63]. Furthermore, fully cocreated trials, where the participant defines what he/she wants to evaluate, might have a higher chance of exerting an actual effect on health behavior change. It would be interesting to embed such trials into models of health behavior change, such as the one by Prochaska et al [64], and think about which elements can map to each stage of precontemplation, contemplation, preparation, action,

maintenance, and termination. Building on this, linking N-of-1 trials further to electronic health records in the future has the potential to connect N-of-1 trials into clinical care and clinical workflows and can further enhance the integration of medical research and clinical practice.

Acknowledgments

This work has received funding from the European Union's Horizon 2020 research and innovation program under the grant agreement 826117 Smart4Health, building a citizen-centered EU–electronic health record exchange for personalized health, and from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation; project 491466077).

Authors' Contributions

AMZ, NS, MM, FP, FH, and DFR developed and implemented *StudyU* and *StudyU Designer* under supervision of SK, TS, SW, and EB. SK, SW, TS, and EB conceived the project. BO and JAE provided critical input regarding design aspects. MD, EG, and MZ provided critical input regarding technical and ethical points. SK, SW, and TS drafted the manuscript. All authors reviewed and commented on the drafts of the manuscript and approved the final version for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary text and figures.
[PDF File (Adobe PDF File), 1984 KB - jmir_v24i7e35884_app1.pdf]

References

1. Spear BB, Heath-Chiozzi M, Huff J. Clinical application of pharmacogenetics. Trends Mol Med 2001 May;7(5):201-204. [doi: 10.1016/s1471-4914(01)01986-4] [Medline: 11325631]
2. Leucht S, Helfer B, Gartlehner G, Davis JM. How effective are common medications: a perspective based on meta-analyses of major drugs. BMC Med 2015 Oct 02;13:253 [FREE Full text] [doi: 10.1186/s12916-015-0494-1] [Medline: 26431961]
3. Eichler H, Abadie E, Breckenridge A, Flamion B, Gustafsson LL, Leufkens H, et al. Bridging the efficacy-effectiveness gap: a regulator's perspective on addressing variability of drug response. Nat Rev Drug Discov 2011 Jul 01;10(7):495-506. [doi: 10.1038/nrd3501] [Medline: 21720406]
4. Lillie EO, Patay B, Diamant J, Issell B, Topol EJ, Schork NJ. The n-of-1 clinical trial: the ultimate strategy for individualizing medicine? Per Med 2011 Mar;8(2):161-173 [FREE Full text] [doi: 10.2217/pme.11.7] [Medline: 21695041]
5. Nikles J, Mitchell G. The Essential Guide to N-of-1 Trials in Health. Amsterdam: Springer; 2015.
6. Wilkinson J, Arnold KF, Murray EJ, van Smeden M, Carr K, Sippy R, et al. Time to reality check the promises of machine learning-powered precision medicine. Lancet Digit Health 2020 Dec;2(12):e677-e680 [FREE Full text] [doi: 10.1016/S2589-7500(20)30200-4] [Medline: 33328030]
7. Bentley KH, Kleiman EM, Elliott G, Huffman JC, Nock MK. Real-time monitoring technology in single-case experimental design research: Opportunities and challenges. Behav Res Ther 2019 Jun;117:87-96. [doi: 10.1016/j.brat.2018.11.017] [Medline: 30579623]
8. Blackston JW, Chapple AG, McGree JM, McDonald S, Nikles J. Comparison of Aggregated N-of-1 Trials with Parallel and Crossover Randomized Controlled Trials Using Simulation Studies. Healthcare (Basel) 2019 Nov 06;7(4):137 [FREE Full text] [doi: 10.3390/healthcare7040137] [Medline: 31698799]
9. Punja S, Schmid CH, Hartling L, Urichuk L, Nikles CJ, Vohra S. To meta-analyze or not to meta-analyze? A combined meta-analysis of N-of-1 trial data with RCT data on amphetamines and methylphenidate for pediatric ADHD. J Clin Epidemiol 2016 Aug;76:76-81. [doi: 10.1016/j.jclinepi.2016.03.021] [Medline: 27060386]
10. Mirza RD, Punja S, Vohra S, Guyatt G. The history and development of N-of-1 trials. J R Soc Med 2017 Aug;110(8):330-340 [FREE Full text] [doi: 10.1177/0141076817721131] [Medline: 28776473]
11. Kravitz RL, Schmid CH, Marois M, Wilsey B, Ward D, Hays RD, et al. Effect of Mobile Device-Supported Single-Patient Multi-crossover Trials on Treatment of Chronic Musculoskeletal Pain: A Randomized Clinical Trial. JAMA Intern Med 2018 Oct 01;178(10):1368-1377 [FREE Full text] [doi: 10.1001/jamainternmed.2018.3981] [Medline: 30193253]
12. Knottnerus JA, Tugwell P, Tricco AC. Individual patients are the primary source and the target of clinical research. J Clin Epidemiol 2016 Aug;76:1-3. [doi: 10.1016/j.jclinepi.2016.07.005] [Medline: 27670675]
13. International Collaborative Network for N-of-1 Clinical Trials and Single-Case Designs. URL: https://www.nof1sced.org/ [accessed 2022-04-17]

14. Müller AR, Brands MMMG, van de Ven PM, Roes KCB, Cornel MC, van Karnebeek CDM, et al. Systematic Review of N-of-1 Studies in Rare Genetic Neurodevelopmental Disorders: The Power of 1. Neurology 2021 Mar 16;96(11):529-540 [FREE Full text] [doi: 10.1212/WNL.0000000000011597] [Medline: 33504638]
15. Porcino AJ, Shamseer L, Chan A, Kravitz RL, Orkin A, Punja S, SPENT group. SPIRIT extension and elaboration for n-of-1 trials: SPENT 2019 checklist. BMJ 2020 Feb 27;368:m122. [doi: 10.1136/bmj.m122] [Medline: 32107202]
16. Vohra S, Shamseer L, Sampson M, Bukutu C, Schmid CH, Tate R, CENT Group. CONSORT extension for reporting N-of-1 trials (CENT) 2015 Statement. BMJ 2015 May 14;350:h1738 [FREE Full text] [doi: 10.1136/bmj.h1738] [Medline: 25976398]
17. Tate RL, Perdices M, Rosenkoetter U, Wakim D, Godbee K, Togher L, et al. Revision of a method quality rating scale for single-case experimental designs and n-of-1 trials: the 15-item Risk of Bias in N-of-1 Trials (RoBiNT) Scale. Neuropsychol Rehabil 2013;23(5):619-638. [doi: 10.1080/09602011.2013.824383] [Medline: 24050810]
18. Stunnenberg BC, Deinum J, Nijenhuis T, Huysmans F, van der Wilt GJ, van Engelen BGM, et al. N-of-1 Trials: Evidence-Based Clinical Care or Medical Research that Requires IRB Approval? A Practical Flowchart Based on an Ethical Framework. Healthcare (Basel) 2020 Feb 27;8(1):49 [FREE Full text] [doi: 10.3390/healthcare8010049] [Medline: 32120865]
19. Product: N Of 1 Health Research. Digital Infuzion. URL: https://www.digitalinfuzion.com/products-patents/n-of-1-health-research/ [accessed 2022-04-17]
20. Zenner A, Böttinger E, Konigorski S. StudyMe: a new mobile app for user-centric N-of-1 Trials. arXiv :2108.00320 Preprint posted online July 31, 2021 [FREE Full text]
21. Daskalova N, Kyi E, Ouyang K, Borem A, Chen S, Park SH, et al. Self-E: Smartphone-Supported Guidance for Customizable Self-Experimentation. 2021 Presented at: CHI '21: CHI Conference on Human Factors in Computing Systems; May 8-13, 2021; Yokohama URL: https://doi.org/10.1145/3411764.3445100 [doi: 10.1145/3411764.3445100]
22. Barr C, Marois M, Sim I, Schmid CH, Wilsey B, Ward D, et al. The PREEMPT study - evaluating smartphone-assisted n-of-1 trials in patients with chronic pain: study protocol for a randomized controlled trial. Trials 2015 Feb 27;16:67 [FREE Full text] [doi: 10.1186/s13063-015-0590-8] [Medline: 25881274]
23. Golden E, Johnson M, Jones M, Viglizzo R, Bobe J, Zimmerman N. Measuring the Effects of Caffeine and L-Theanine on Cognitive Performance: A Protocol for Self-Directed, Mobile N-of-1 Studies. Front Comput Sci 2020 Feb 13;2:4. [doi: 10.3389/fcomp.2020.00004]
24. Bot BM, Suver C, Neto EC, Kellen M, Klein A, Bare C, et al. The mPower study, Parkinson disease mobile data collected using ResearchKit. Sci Data 2016 Mar 03;3:160011 [FREE Full text] [doi: 10.1038/sdata.2016.11] [Medline: 26938265]
25. Ngari MM, Waithira N, Chilengi R, Njuguna P, Lang T, Fegan G. Experience of using an open source clinical trials data management software system in Kenya. BMC Res Notes 2014 Nov 26;7:845 [FREE Full text] [doi: 10.1186/1756-0500-7-845] [Medline: 25424974]
26. Taylor S, Sano A, Ferguson C, Mohan A, Picard RW. QuantifyMe: An Open-Source Automated Single-Case Experimental Design Platform. Sensors (Basel) 2018 Apr 05;18(4):1097 [FREE Full text] [doi: 10.3390/s18041097] [Medline: 29621133]
27. Karkar R, Schroeder J, Epstein DA, Pina LR, Scofield J, Fogarty J, et al. TummyTrials: A Feasibility Study of Using Self-Experimentation to Detect Individualized Food Triggers. Proc SIGCHI Conf Hum Factor Comput Syst 2017 May 02;2017:6850-6863 [FREE Full text] [doi: 10.1145/3025453.3025480] [Medline: 28516175]
28. Daskalova N, Metaxa D, Tran A, Nugent NR, Boergers J, McGeary J, et al. SleepCoacher: A Personalized Automated Self-Experimentation System for Sleep Recommendations. 2016 Presented at: UIST '16: The 29th Annual ACM Symposium on User Interface Software and Technology; October 16-19, 2016; Tokyo. [doi: 10.1145/2984511.2984534]
29. Evans B. Paco-Applying Computational Methods to Scale Qualitative Methods. Ethnogr Prax Ind Conf Proc 2016 Nov 29;2016(1):348-368. [doi: 10.1111/1559-8918.2016.01095]
30. Giessing L, Oudejans RRD, Hutter V, Plessner H, Strahler J, Frenkel MO. Acute and Chronic Stress in Daily Police Service: A Three-Week N-of-1 Study. Psychoneuroendocrinology 2020 Dec;122:104865. [doi: 10.1016/j.psyneuen.2020.104865] [Medline: 32961407]
31. Mühlbauer E, Bauer M, Ebner-Priemer U, Ritter P, Hill H, Beier F, et al. Effectiveness of smartphone-based ambulatory assessment (SBAA-BD) including a predicting system for upcoming episodes in the long-term treatment of patients with bipolar disorders: study protocol for a randomized controlled single-blind trial. BMC Psychiatry 2018 Oct 26;18(1):349 [FREE Full text] [doi: 10.1186/s12888-018-1929-y] [Medline: 30367608]
32. mPower. GitHub. URL: https://github.com/Sage-Bionetworks/mPower [accessed 2022-04-17]
33. Parkinson mPower 2. Apple App Store. URL: https://apps.apple.com/us/app/parkinson-mpower-2/id1375781575 [accessed 2022-04-17]
34. PACO: The Personal Analytics Companion. URL: https://pacoapp.com [accessed 2022-04-17]
35. movisensXD. movisens GmbH. URL: https://www.movisens.com/en/products/movisensxs/ [accessed 2022-04-17]
36. OpenClinica. URL: https://www.openclinica.com [accessed 2022-04-17]
37. QuantifyMe. GitHub. URL: https://github.com/mitmedialab/AffectiveComputingQuantifyMeAndroid [accessed 2022-04-17]
38. TummyTrials. GitHub. URL: https://github.com/tractdb/tummytrials [accessed 2022-04-17]
39. SleepCoacher. URL: https://sleepcoacher.cs.brown.edu/ [accessed 2022-04-17]
40. StudyU platform. GitHub. URL: https://github.com/hpi-studyu [accessed 2022-04-17]

41. StudyU Health app. Google Play. URL: https://play.google.com/store/apps/details?id=health.studyu.app [accessed 2022-04-17]
42. StudyU Health app. Apple App Store. URL: https://apps.apple.com/us/app/studyu-health/id1571991198 [accessed 2022-04-17]
43. Flutter. URL: https://flutter.dev [accessed 2022-04-17]
44. Parse. URL: https://parseplatform.org [accessed 2022-04-17]
45. Docker. URL: https://www.docker.com [accessed 2022-04-17]
46. Back4App. URL: https://www.back4app.com [accessed 2022-04-17]
47. StudyU designer. URL: https://studyu-designer-v1.web.app [accessed 2022-04-17]
48. StudyU app. URL: https://studyu-app-v1.web.app/#/welcome [accessed 2022-04-17]
49. The European Parliament and the Council of the European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). Official Journal of the European Union 2016 [FREE Full text]
50. Benabou R, Tirole J. Intrinsic and Extrinsic Motivation. Rev Econ Studies 2003 Jul;70(3):489-520 [FREE Full text] [doi: 10.1111/1467-937x.00253]
51. Villar A, Callegaro M, Yang Y. Where Am I? A Meta-Analysis of Experiments on the Effects of Progress Indicators for Web Surveys. Soc Sci Comput Rev 2013 Aug 19;31(6):744-762 [FREE Full text] [doi: 10.1177/0894439313497468]
52. Conrad FG, Couper MP, Tourangeau R, Peytchev A. The impact of progress indicators on task completion. Interact Comput 2010 Sep 01;22(5):417-427 [FREE Full text] [doi: 10.1016/j.intcom.2010.03.001] [Medline: 20676386]
53. Nunes J, Drèze X. The Endowed Progress Effect: How Artificial Advancement Increases Effort. J CONSUM RES 2006 Mar;32(4):504-512 [FREE Full text] [doi: 10.1086/500480]
54. Bica I, Alaa AM, Lambert C, van der Schaar M. From Real-World Patient Data to Individualized Treatment Effects Using Machine Learning: Current and Future Methods to Address Underlying Challenges. Clin Pharmacol Ther 2021 Jan;109(1):87-100. [doi: 10.1002/cpt.1907] [Medline: 32449163]
55. Shalit U, Johansson F, Sontag D. Estimating individual treatment effect: generalization bounds and algorithms. 2017 Presented at: International Conference on Machine Learning. PMLR; 2017; Sydney, Australia p. 3076-3085 URL: https://proceedings.mlr.press/v70/shalit17a/shalit17a.pdf
56. Lee C, Mastronarde N, van DSM. Estimation of Individual Treatment Effect in Latent Confounder Models via Adversarial Learning. arXiv: 1811.08943 2018 [FREE Full text]
57. Alaa A, van DSM. Bayesian Inference of Individualized Treatment Effects using Multi-task Gaussian Processes. arXiv Preprint posted online May 28, 2017 [FREE Full text]
58. Boruvka A, Almirall D, Witkiewitz K, Murphy SA. Assessing Time-Varying Causal Effect Moderation in Mobile Health. J Am Stat Assoc 2018;113(523):1112-1121 [FREE Full text] [doi: 10.1080/01621459.2017.1305274] [Medline: 30467446]
59. Walton AE, Collins LM, Klasnja P, Nahum-Shani I, Rabbi M, Walton MA, et al. The Micro-Randomized Trial for Developing Digital Interventions: Experimental Design Considerations. arXiv Preprint posted online April 23, 2020 [FREE Full text]
60. Qian T, Russell MA, Collins LM, Klasnja P, Lanza ST, Yoo H, et al. The micro-randomized trial for developing digital interventions: data analysis methods. arXiv Preprint posted online April 21, 2020 [FREE Full text]
61. 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 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
62. Chen X, Chen P. A comparison of four methods for the analysis of N-of-1 trials. PLoS One 2014;9(2):e87752 [FREE Full text] [doi: 10.1371/journal.pone.0087752] [Medline: 24503561]
63. Karkar R, Zia J, Vilardaga R, Mishra SR, Fogarty J, Munson SA, et al. A framework for self-experimentation in personalized health. J Am Med Inform Assoc 2016 May;23(3):440-448 [FREE Full text] [doi: 10.1093/jamia/ocv150] [Medline: 26644399]
64. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]

Abbreviations

CSV: comma-separated values
GDPR: General Data Protection Regulation
HIPAA: Health Insurance Portability and Accountability Act
IRB: institutional review board
JSON: JavaScript Object Notation
RCT: randomized controlled trial
SCED: single-case experimental design

Edited by T Leung; submitted 21.12.21; peer-reviewed by J Nikles, S Mangelsdorf; comments to author 16.03.22; revised version received 17.04.22; accepted 18.04.22; published 05.07.22.

Please cite as:
Konigorski S, Wernicke S, Slosarek T, Zenner AM, Strelow N, Ruether DF, Henschel F, Manaswini M, Pottbäcker F, Edelman JA, Owoyele B, Danieletto M, Golden E, Zweig M, Nadkarni GN, Böttinger E
StudyU: A Platform for Designing and Conducting Innovative Digital N-of-1 Trials
J Med Internet Res 2022;24(7):e35884
URL: https://www.jmir.org/2022/7/e35884
doi:10.2196/35884
PMID:35787512

Viewpoint

Assessing a New Prescreening Score for the Simplified Evaluation of the Clinical Quality and Relevance of eHealth Apps: Instrument Validation Study

Nicolas Wagneur[1], MSc; Patrick Callier[2,3], PharmD, PhD; Jean-David Zeitoun[3,4], MD, PhD; Denise Silber[3,5], MBA; Remi Sabatier[3,6], MD; Fabrice Denis[1,3], MD, PhD

[1]Institut Inter-régional Jean Bernard, ELSAN, Le Mans, France

[2]Laboratoire de génétique chromosomique et moléculaire, Centre Hospitalier Universitaire, Dijon, France

[3]Institut National de la e-Santé, Le Mans, France

[4]Centre d'Epidémiologie Clinique, Hôtel Dieu Hospital, Assistance Publique-Hopitaux de Paris, Paris, France

[5]Basil Strategies, Paris, France

[6]Service de Cardiologie, Centre Hospitalier Universitaire de Caen, Caen, France

Corresponding Author:
Fabrice Denis, MD, PhD
Institut National de la e-Santé
9 rue Beauverger
Le Mans, 72000
France
Phone: 33 684190480
Email: f.denis@ilcgroupe.fr

Abstract

Background: In 2020, more than 250 eHealth solutions were added to app stores each day, or 90,000 in the year; however, the vast majority of these solutions have not undergone clinical validation, their quality is unknown, and the user does not know if they are effective and safe. We sought to develop a simple prescreening scoring method that would assess the quality and clinical relevance of each app. We designed this tool with 3 health care stakeholder groups in mind: eHealth solution designers seeking to evaluate a potential competitor or their own tool, investors considering a fundraising candidate, and a hospital clinician or IT department wishing to evaluate a current or potential eHealth solution.

Objective: We built and tested a novel prescreening scoring tool (the Medical Digital Solution scoring tool). The tool, which consists of 26 questions that enable the quick assessment and comparison of the clinical relevance and quality of eHealth apps, was tested on 68 eHealth solutions.

Methods: The Medical Digital Solution scoring tool is based on the 2021 evaluation criteria of the French National Health Authority, the 2022 European Society of Medical Oncology recommendations, and other provided scores. We built the scoring tool with patient association and eHealth experts and submitted it to eHealth app creators, who evaluated their apps via the web-based form in January 2022. After completing the evaluation criteria, their apps obtained an overall score and 4 categories of subscores. These criteria evaluated the type of solution and domain, the solution's targeted population size, the level of clinical assessment, and information about the provider.

Results: In total, 68 eHealth solutions were evaluated with the scoring tool. Oncology apps (22%, 20/90) and general health solutions (23%, 21/90) were the most represented. Of the 68 apps, 32 (47%) were involved in remote monitoring by health professionals. Regarding clinical outcomes, 5% (9/169) of the apps assessed overall survival. Randomized studies had been conducted for 21% (23/110) of the apps to assess their benefit. Of the 68 providers, 38 (56%) declared the objective of obtaining reimbursement, and 7 (18%) out of the 38 solutions seeking reimbursement were assessed as having a high probability of reimbursement. The median global score was 11.2 (range 4.7-17.4) out of 20 and the distribution of the scores followed a normal distribution pattern (Shapiro-Wilk test: P=.33).

Conclusions: This multidomain prescreening scoring tool is simple, fast, and can be deployed on a large scale to initiate an assessment of the clinical relevance and quality of a clinical eHealth app. This simple tool can help a decision-maker determine which aspects of the app require further analysis and improvement.

(J Med Internet Res 2022;24(7):e39590) doi:10.2196/39590

KEYWORDS

scoring; eHealth; clinical relevance; solution; digital solution; clinical validation; prescreening; eHealth app; medical digital solution; scoring tool; health app; information quality

Introduction

The number of eHealth tools has been expanding with the acceleration of innovation in telemedicine, connected objects, artificial intelligence, electronic patient-reported outcomes, immersive technologies, and other fields.

The COVID-19 pandemic further accelerated the emergence of new eHealth apps [1-3]. In 2020, 327,000 health apps were available on the Android and iOS App Store, and more than 250 eHealth solutions were added to app stores each day, or 90,000 in the year [4]. The number of health apps had doubled since 2013 [5].

However, there is great heterogeneity in the quality, relevance, and clinical performance of these solutions. It is difficult for users to differentiate the apps according to these 3 major criteria. It is also challenging for the providers of the eHealth apps to comply with good clinical practice. The technical developers may have no medical background or access to practicing clinicians. Most apps propose imprecise clinical benefits, and since they do not undergo any regulatory processes, their quality is uncertain and variable.

Whereas existing evaluation scores are often complex and difficult to deploy, health care institutions and the prescribers of these eHealth solutions need a simple, quick prescreening tool. However, there is no consensus on the benchmark for evaluating them in the context of clinical activity.

These prescreening tools must be based on good clinical practice guides and recommendations. Many standards and scoring methods already exist, and the first international recommendations for remote monitoring in oncology are now available [6].

A recent review of relevant medical literature analyzed the quality criteria for evaluating health solutions. Other criteria were then provided by the French National Health Authority (Haute Autorité de Santé; HAS), which is also responsible for the evaluation of drugs and medical devices [7,8]. Various other scores were also identified [9], such as mobile health evidence reporting and assessment [10], Digital Technology Assessment Criteria [11], ORCHA Review score [4], and MyHealthApps [12].

However, these scoring tools may include more than 150 questions, are laborious to use, and their effectiveness is yet unknown. Furthermore, they rarely evaluate all 4 key characteristics: clinical relevance, use potential, the quality of the provider, and the specificities of the solution.

We therefore set out to propose a rapid prescreening evaluation score. Although it can be used by any health care stakeholder, we determined 3 priority target users: eHealth solution designers, potential investors, and hospital decision-makers wishing to evaluate an existing or potential future solution. We developed the scoring tool to assess all aspects of eHealth good clinical practices and evaluated the key categories for 68 digital eHealth solutions.

Methods

Medical Digital Solution Scoring Tool

We built and made available the Medical Digital Solution (MDS) scoring tool, a new prescreening scoring tool based on 26 questions. We then used this tool to evaluate a panel of eHealth solutions [13].

The solution frontend was programmed with the ReactJS language. The application is hosted on a Hostinger server secured by an SSL protocol. The backend is based on the NoSQL Firebase solution. The technical functionality of the electronic questionnaire was tested by 10 editors before fielding the questionnaire. No cookies were used, and no IP check was done.

The design of the MDS scoring tool was based on the 2021 HAS Solution Evaluation Criteria [7], the HAS Good Practices Framework on Solutions and Connected Objects in Health (eHealth or mobile Health) of 2016 [14], and the European Society of Medical Oncology (ESMO) recommendations of 2022 [6].

This evaluation score was presented to the providers of eHealth solutions via a campaign on LinkedIn (a professional social network) from January 18, 2022, to January 30, 2022. The survey announcement is detailed in Multimedia Appendix 1, and the following is an English-translated excerpt:

> *How can we quickly assess the relevance and potential of a medical digital solution? We created the MDS trust score which aims to provide a rapid assessment of digital medical solutions tool for software publishers, patient associations, investors in the field of eHealth, and institutions. It is available to startups/solution publishers, associations, institutions, and investors...If you are interested in using it, please contact us.*

The evaluation of the eHealth solutions was conducted via a close-access web solution URL [13]. The solutions retained for evaluation had to have clinical objectives. Wellness solutions were excluded, and we kept only the first evaluation to limit false score optimization biases. Only completed questionnaires were analyzed.

Solution providers were examined in light of the evaluation criteria and given a score in each of the 4 categories, as well as a total score. The categories included the specificities of the solution, the solution's targeted population size and use

potential, the clinical evaluation information of the solution, and provider information.

Part 1—solution specificities—evaluated the scope of the solution, the specialty concerned by the solution, the type of solution used, its compliance with the digital doctrine established by the HAS (a public agency reporting to the French Ministry of Health), the type of algorithm used by the solution, as well as its capacity of interacting with the user [15].

Part 2 assessed the solution's target users based on age, user volume in France, the possibility for the use of the tool outside of France, the degree of its codevelopment with patients or patient associations, and the impact of the solution on the hospital organization.

Part 3 focused on the clinical evaluation of the solution, the outcomes used to assess the clinical benefit of the solution, the inclusion of feedback from medical specialists regarding the clinical relevance of the solution, the presence of support from or partnership with a scientific society, as well as the current level of clinical evidence of the solution.

Part 4 focused on the evaluation of the provider and included the presence of fundraising, the country of the headquarters, the presence of founding doctors on the board of directors, the presence of a medical department led by a physician, the presence and composition of a medical and scientific board, the media visibility of the solution on the internet, the development of previous eHealth solutions by the provider, and the presence or absence of a critical safety alert from the French National Agency for the Safety of Medicines. The strength and goals of the business model and reimbursement by French social security were also assessed. Among providers seeking reimbursement, we calculated a reimbursement probability score based on the clinical evaluation of the solution. The result was expressed in the semiquantitative form (low, medium, or high probability of reimbursement).

A score out of 500 was assigned to each of the 4 categories resulting in an overall score out of 2000, which was then reduced to a score out of 20. An example of the MDS tool is shown in Multimedia Appendix 2.

The questions and the weighting of the different answers to the 26 questions were designed by a group of 16 medical experts, eHealth experts, representatives of manufacturers and eHealth solution providers, methodologists, institutional evaluators, eHealth researchers, and representatives of patient associations.

A tool within the web platform also allowed providers to rank their solutions against other tested solutions.

Ethical Considerations

No ethics review board assessment was required for this study of the characteristics of the solutions given the absence of patient data analysis and intervention. No demographic data were available, and their collection would not have been appropriate, as we only assessed solution characteristics and not the health data the solutions would collect.

Statistical Analysis

We carried out a descriptive study of the characteristics of the solutions and assessed the scores of each solution by calculating the average, the median, and the first and third quartile distribution of the solutions. A Shapiro-Wilk test was performed to determine if the distribution of the score followed a normal distribution. For the chosen alpha level of .05, the scores were considered as normally distributed if the *P* value was >.05.

Results

The MDS assessment score was used for 135 eHealth solutions via the web solution, and 68 solutions were assessable for our analysis. For the other solutions, the data were either incomplete (n=17) or duplicate (n=50). Incomplete forms were excluded from the analysis.

The 68 assessable solutions were associated with 102 clinical indications. Of the 102 clinical indications, 31 (30%) were related to support in taking medications, medicine compliance, and the reduction of treatment toxicity; 23 (23%) concerned the early detection of disease; 16 (16%) were related to decision support; 12 (12%) concerned prevention; 6 (6%) concerned direct therapeutic indications; 2 (2%) were related to patient triage; and 2 (2%) were aimed at relieving emergency department overload.

Of the 68 solutions, 22 (32%) targeted several medical specialties. Of the 90 specialties, the most present specialties were oncology with 20 (22%) solutions and cross-cutting solutions such as pain management with 21 (23%) solutions.

Of the 68 evaluable solutions, 29 (43%) were Class I medical devices according to the Medical Devices Directive 93/42/EEC of European Union [16] (Table 1).

Part 1 of the score concerned the study of general information about the solution (Table 2). Of the 68 evaluable solutions, 28 (41%) were based on nonartificial intelligence algorithms, and 22 (32%) were based on algorithms using artificial intelligence, of which 6 (9%) contained a nonintelligible artificial intelligence algorithm. We noted that almost all the algorithms (n=67, 99%) were less than 5 years old or otherwise up to date regarding the clinical standards within their specialty. Of the 68 solutions studied, 65 (96%) had a user interaction system and 32 (47%) were associated with remote monitoring with a health care professional.

Table 1. Solution characteristics (several items were possible per solution).

Characteristic	Solution, n (%)
Solution area (2 items maximum per solution; N=102)	
Prevention	12 (12)
Early detection and flagging	23 (23)
Decision support	16 (16)
Treatment support, compliance, and toxicity reduction	31 (31)
Direct therapeutic solution (eg, virtual reality)	6 (6)
Patient triage	2 (2)
Emergency department decongestion	2 (2)
Others	6 (6)
Specialty concerned by the solution (2 items maximum per solution; N=90)	
Oncology	20 (22)
Cardiology	5 (6)
Neurology	6 (7)
Psychiatry	5 (6)
Pediatrics	2 (2)
Diabetology	7 (8)
Gynecology	5 (6)
Pulmonology	1 (1)
Nephrology	1 (1)
Urology and andrology	1 (1)
Rheumatology	2 (2)
Head and neck	1 (1)
Gastroenterology	0 (0)
Dermatology	3 (3)
Autoimmune disease, internal medicine, and infectiology	1 (1)
Surgery	3 (3)
Imaging	2 (2)
Geriatrics	3 (3)
Nutrition	1 (1)
Ophthalmology	0 (0)
Multiple specialties (emergency medicine, general medicine, and biology, etc)	21 (23)
Solution type (N=68)	
Nonmedical device (nonexecutive)	15 (22)
Nonmedical device (but theoretically should be)	10 (15)
Class I medical device	29 (43)
Class II medical device	12 (18)
Other CE[a] markings (eg, Class III medical device)	2 (3)

[a]CE: Conformitè Europëenne.

Table 2. Part 1: general information about the solutions.

Characteristic	Solution (N=68), n (%)
Type of algorithm used in the solution	
No artificial intelligence (Boolean, common rules, and logistic regression, etc)	28 (41)
No algorithm	17 (25)
Intelligible artificial intelligence	16 (24)
Nonintelligible artificial intelligence	6 (9)
Algorithm more than 5 years old that were not reassessed with new support standards	1 (1)
Possibility of interaction with the user	
Yes, with remote monitoring with a health professional	32 (47)
Yes, alerts to the patient who then manages themselves	15 (22)
Yes, information not personalized according to the answers	10 (15)
Yes, patient alert with teleconsultation possible via the solution	3 (4)
Solution not affected	5 (7)
None, no user information	3 (4)

Part 2 of the score concerned the study of the target population of the solution (Table 3). Of the 68 solutions evaluated, 42 (62%) covered several population age groups. The main age groups of the target populations were adults aged 18-64 years (58/110, 53% of age groups) and people aged ≥65 years (42/110, 38% of age groups). The median size of the population potentially reached by the solutions in France was 100,000 people. Of the 68 evaluable solutions, 59 (87%) were potentially applicable to the rest of the world. Patients had been involved in the solution development process in 59 (87%) solutions. The solutions tested typically facilitated the simplification of hospital organization (59%, n=40).

Part 3 of the score assessed the level of evidence and clinical relevance of the solution (Table 4). The clinical outcomes evaluated by the providers were heterogeneous and often multiple. In our study, we did not find a single, common criterion. Of the 169 validated outcomes, user satisfaction was cited 29 (17%) times, quality of life 24 (14%) times, medico-economic benefit 20 (12%) times, gain in early diagnosis 19 (11%) times, improved treatment compliance 19 (11%) times, and overall survival 9 (5%) times. Of the 68 solutions, 47 (69%) were assessed by experts as having major relevance, and 38 (56%) providers had benefited from the support of or partnership with a scientific society. Regarding the level of evidence of the solutions, 110 clinical evaluations were conducted for these solutions; 23 (21%) were randomized, 17 (15%) were prospective nonrandomized studies, 18 (16%) were retrospective studies, 28 (26%) were based on expert agreement, and 8 (7%) were not based on any studies or expert opinions.

Table 3. Part 2: the target populations of the solutions.

Characteristic	Solution, n (%)
Age of targeted population (years; N=110)	
<18 (pediatrics)	10 (9)
18-64	58 (53)
>65	42 (38)
Number of patients involved in the codevelopment of the solution (N=68)	
>500	20 (29)
50-499	24 (35)
1-49	15 (22)
0	9 (13)
Impact on hospital organization (N=68)	
Simplification	40 (59)
Complication without associated act or package	7 (10)
Complication with associated act or package for financing the organization	4 (6)
No impact	17 (25)

Table 4. Part 3: the clinical relevance of the solutions.

Characteristic	Solution, n (%)
Validated outcome (several possible items per solution; N=169)	
User satisfaction	29 (17)
Quality of life	24 (14)
Medico-economic benefit	20 (12)
Early diagnosis gain	19 (11)
Improved access to care	19 (11)
Improved treatment compliance	18 (11)
Reduction of severity of a condition or symptom	13 (8)
Primary, secondary, or tertiary prevention	11 (7)
Survival	9 (5)
Reduction of emergencies	4 (2)
Less toxicity than reference	3 (2)
Level of evidence based on clinical assessment (several possible items per solution; N=110)	
Expert advice	28 (25)
Retrospective study of ≥300 evaluable patients	18 (16)
Applications of national or international recommendations in the solution	16 (15)
Randomized trial of <200 patients	13 (12)
Randomized trial of ≥200 patients or a meta-analysis	10 (9)
Prospective study of ≥200 patients versus nonrandomized control arm in real-life settings (historical comparison and data, etc)	11 (10)
Prospective study of ≥200 patients -) versus nonrandomized control arm not in real-life settings	6 (5)
Not based on any studies or expert opinions	8 (7)

Part 4 the score concerned the characteristics of the provider (Table 5). Regarding provider fundraising, the answer was provided by 60 (88%) out of the 68 providers. Of these 60 providers, 30 (50%) had not yet raised funds, 11 (18%) had raised between €1.5 million (US$ 1.59 million) and €5 million (US $5.28 million), and 11 (18%) had raised >€5 million (>US $5.28 million). Of the 68 providers, 55 (81%) were based in France, 10 (15%) in other European countries, 1 (2%) in the United States, and 2 (3%) outside of Europe and the United States; 51 (75%) had a medical department that included at least one physician, of which 41 (60%) included at least one specialist in the field of the solution; and 60 (96%) had a scientific board with at least one doctor.

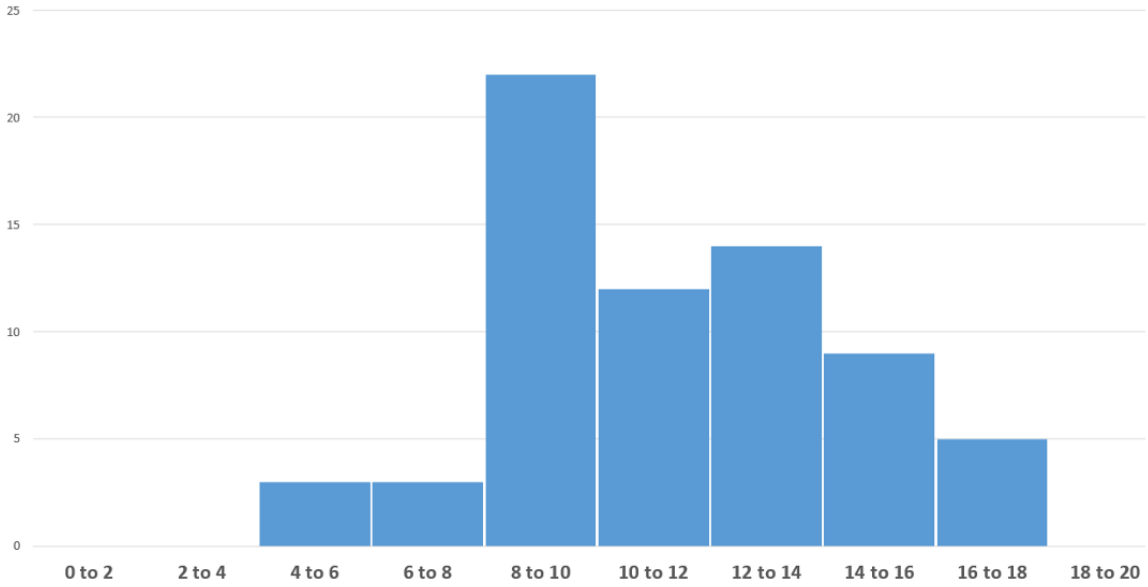
The media awareness of the solution over the past 12 months was assessed. On average, the providers or their solution were listed in 12 Google News search results, with a median of 7 search results. Of the 68 providers, 37 (54%) were developing their first eHealth app. Regarding the security of the solutions, 9 (13%) providers had had a security alert from the French National Agency for the Safety of Medicines. Of the 68 providers, 38 (56%) intended to obtain social security reimbursement for their solution. Of these 38 solutions, 26 (68%) had a low probability of reimbursement and 7 (18%) had a high probability.

The calculation of the overall score is carried out for each eHealth solution by summing the points of the 4 previous criteria (Figure 1). The average score was 11.25 (range 4.7-17.4) points out of 20, the median score was 11.2 points out of 20, and the distribution followed a normal distribution (Shapiro-Wilk test: P=.33). The top 25% of apps scored below 9.4 out of 20, whereas the top 25% of apps scored above 13.4 out of 20.

Table 5. Part 4: information on the solution provider and reimbursement ambition and probability.

Characteristic	Solution, n (%)
Presence of a medical director (N=68)	
Medical management by a specialist doctor	41 (60)
Medical management by a nonspecialist doctor	10 (15)
No medical direction	17 (25)
Development of previous eHealth solutions or medical device with >500 users (N=68)	
Yes, 3 or more	11 (16)
Yes, 2	7 (10)
Yes, 1	13 (19)
No	37 (54)
Business model (ambition or reimbursement as a goal; N=68)	
Yes (unless “device for collective hospital use”)	38 (56)
No, no need in the business model (sale of data and user subscription, etc)	23 (34)
No, device for collective hospital use	7 (10)
Probability of reimbursement (if an objective of the solution; n=38)	
High	7 (18)
Medium	5 (13)
Low	26 (68)

Figure 1. Overall rating distribution of the solutions according to overall score out of 20.



Discussion

Principal Findings

We developed the first multidomain prescreening scoring tool to initiate an assessment of the clinical relevance and quality of a clinical eHealth app.

We proceeded to a first assessment of the relevance, quality, and level of evidence of an eHealth solution for 68 eHealth solutions available in France and reported the characteristics of the solutions in the different assessed fields.

The most represented medical fields were oncology (22%) and cross-cutting solutions covering several specialties (23%). This is confirmed by the literature; the specialty areas that have the most clinically validated eHealth solutions in terms of quality of life or survival are oncology and cardiology [17,18].

In our study, almost half (47%) of the evaluated solutions were based on a remote monitoring system deployed with a health professional. This type of solution occupies an important place in eHealth and is frequently used in cardiology and oncology. The HAS reported this as one of the most common configurations in its 2021 report evaluating solutions in the health sector [7]. The first international recommendations further

stimulated the development of this type of instrument. These recommendations include the quality criteria to consider for the choice of these tools: the level of clinical evidence, the type of algorithm, Conformitè Europëenne marking, and the characteristics of the algorithms.

Evaluating the level of clinical evidence of a solution is an important step for the acceleration of its use in the medical world and possible reimbursement by health authorities. In 2022, this evaluation was a major criterion for obtaining a favorable recommendation from the ESMO for use [6]. We noted a great heterogeneity in the clinical evaluation criteria of the solutions. Only 5% of the solutions used overall survival as an endpoint. This is both one of the most difficult outcomes to obtain and the most relevant criterion according to the scientific community. Several remote monitoring solutions reduce mortality in patients followed for oncological or cardiac pathologies [19,20]. Survival is not an applicable outcome for the majority of the eHealth instruments. The criteria most frequently reported in our study are quality of life, gain in early diagnosis, better medico-economic benefit, or improved compliance with treatment. These criteria remain of interest in many solutions for the patients concerned.

In addition, in our study, 21% of the solutions were the subject of a randomized study. This type of study is considered as the highest degree of evidence and a major criterion to obtain reimbursement in France when undertaken. We also observed that 15% of the solutions had conducted prospective nonrandomized studies. The different types of study were weighted differently in our score. For example, conducting a randomized clinical study of ≥200 patients contributed 190 points out of 500 in the clinical evaluation score, whereas a retrospective study provided only 30 points.

About half (56%) of the providers declared the objective of obtaining reimbursement from French National Public Insurance, which in France covers the totality of the population by law. This possibility has been available in France since 2018. It adds a new business model modality to the development of a solution. We used the evaluation tool to identify the solutions that would have a high probability of reimbursement from the French National Social Security. This assessment was based on the type of studies conducted, as well as the type of clinical endpoints measured. In our study, 18% of the solutions among those seeking reimbursement were assessed as having a high probability of reimbursement.

Overall, the results of the evaluation of the 68 eHealth solutions seem close to the known elements of the literature. The average prescreening score of the evaluated solutions was 11.25 points out of 20. The scores ranged from 4.7 to 17.4. The distribution of the scores followed a normal distribution.

Comparison to Prior Work

The first scores for evaluating eHealth solutions that appeared in the literature were mainly based on user or expert opinions [21]. The HAS listed 7 scores that focus on this scope (MyHealthApps [12], GGD Appstore [22], Health Navigator [23], One Mind [24], Osservatorio APP sanitarie [25], HealthOn [26], and the mobile app rating scale [27]). This type of evaluation is important in the development of a solution. In our study, 87% of the solutions involved patients in the development process, 75% had a medical department composed of at least one physician, and 96% had a scientific board with at least one doctor. These elements are important to optimize medical quality and therefore the trust and acceptability by patients of eHealth solutions.

The evaluation frameworks of other scores are typically descriptive, time-consuming, and qualitative tools to assess clinical quality. The Digital Technology Assessment Criteria designed by the National Health Service in the United Kingdom is an example of such an evaluation framework [11].

Strengths and Limitations

We propose the calculation of a score based on our set of 26 evaluation criteria selected by a panel of experts and recommended in the literature. This score simultaneously evaluates the information on the solution, the target of the solution, the clinical evaluation of the solution, and the provider. This prescreening score has the advantage of being quick to achieve and having a wide range of evaluated criteria. The simple evaluation scores proposed in the literature often do not allow an evaluation that is as broad as our score's [7]. Our questions are easily understandable and verifiable. This score can therefore be used by health care professionals including physician prescribers, pharmacists, and nurses; patient associations; investors; and providers to compare their solutions against competitors and track improvements of their solutions. The rapid realization of this score allows it to be regularly recalculated in real time for the same solution to improve the quality of the solutions.

Our score was assessed and validated in 68 eHealth solutions, unlike many other scores proposed in the literature that were not assessed in real life [7]. Notably, one of the most used scores today is the E-Solution Rating Scale [11].

Our prescreening score was developed based on recent and updated recommendations. It is based, among other things, on the recommendations of the HAS guides [7], as well as the recommendations of the ESMO released in 2022 [6]. However, the values of the different parameters will evolve according to new standards, recommendations, and data from the literature.

The limits of this score must also be taken into consideration. First, the score does not allow for an exhaustive and detailed assessment of all technical and clinical criteria. For example, the use of more detailed scores could be used to complete the assessment, such as the ORCHA Review score [28], which evaluates from 260 to 350 criteria; Enlight [29], which evaluates 476 criteria; and the framework from Henson et al [30], which evaluates 357 criteria. Second, the weighting of each answer was discussed by experts but empirically fixed in the absence of applicable quantitative benchmarks. Third, the filling in of the data was done directly by the solution's providers in an autonomous and declarative way. This information was not verified. However, we excluded from the study providers who had not exhaustively filled in the entire questionnaire and duplicate providers when several questionnaires were completed for the same solution—always keeping only the first evaluation

to limit false score optimization biases. Fourth, our tool does not conduct an in-depth assessment of the methodological quality of the reported studies. Other important characteristics of eHealth solutions are outside the scope of this score, including interoperability, security, portability, privacy, regulatory, ethics, and environment. Fifth, the short recruitment of the solution provider sample and the use of LinkedIn as a source can introduce a selection bias of the participants. For example, more sleep or mental health apps could have been assessed with a wider range of recruitment. Moreover, this recruitment led to only evaluating French solutions. Sixth, reimbursement processes are country-dependent, and the only geographical scope considered in our paper is France. The development of an international version of the score is in progress.

Conclusion

We propose a multidomain prescreening tool that is simple and fast to use and usable on a large scale to initiate the evaluation of clinical digital solutions by any health care stakeholder. We believe that 3 target groups (eHealth solution designers, investors, and hospital decision-makers) will be the main initial users. This tool can help improve the quality of solutions and identify the aspects of the tools that may require further analysis and improvement. The score will be accessible on the website on the French National eHealth Institute [31] for the solution providers.

Acknowledgments

We thank the users for their participation in this study as well as Magali Balavoine, MSc, and Anne-Lise Septans from Weprom in Angers, France. We also thank Jean-François Thébault, MD (French Federation of Diabetics); Armelle Graciet (Syndicat National de l'Industrie des Technologies Médicales; the federation representing medical device and “digital” medical device companies); Corinne Collignon (Haute Autorité de Santé); Myriam Oukkal (French eHealth council); Jean-Louis Fraysse, MD; Yann-Mael le Douarin, MD; Alain Toledano, MD; Jean-Jacques Zambrowski, Prof, MD, PhD; and Celine Riou (Banque Publique d’Investissement) for their tips and advice. No one received compensation for their contributions. No funding was necessary to conduct the trial.

Data Availability

The data sets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' Contributions

FD had full access to all of the data in the study, takes responsibility for the integrity of the data and the accuracy of the data analysis, contributed to the concept and design of this study, conducted the statistical analysis, and is the guarantor of the paper. All authors contributed to the acquisition, analysis, or interpretation of data; the drafting of the manuscript; and the critical revision of the manuscript for important intellectual content. FD and RS supervised this study.

Conflicts of Interest

NW and PC declare no conflicts of interest. JDZ reports being an advisor for several consulting firms in link with pharmaceutical industry (Oliver Wyman, Roland Berger). He also reports speaking fees from a manufacturers’ professional association and consulting fees from Ferring, Pierre Fabre, AbbVie, Astra Zeneca, Biogen, Boehringer Ingelheim, Takeda, and Johnson & Johnson. He is a personal investor in approximately 30 digital companies, medical device companies, or biotech companies, and a limited partner in 3 investment funds. He reports being a founding partner of Inato, a company involved in clinical research and whose customers are pharmaceutical companies. DS reports conflicts of interest with Basil Strategies. RS reports conflicts of interests with Amgen, Astra-Zeneca, Biosensors, Bristol-Myers-Squibb, Boston Scientific, Novartis, Robocath, Terumo, and Vifor. FD reports conflicts of interests with Chugai, Astra-Zeneca, Merck, Sivan, Takeda, Ipsen, Bristol Meyer Squibb, Viatris, Kelindi, and Hyperion where he is an invited speaker, and also serves on the advisory board at Sivan and Roche. The author holds stocks and shares with INeS, Kelindi, Hyperion and the institution receiver was Hvperion.

Multimedia Appendix 1
The published survey announcement.
[PNG File , 206 KB - jmir_v24i7e39590_app1.png]

Multimedia Appendix 2
Screenshot of the developed tool.
[PNG File , 118 KB - jmir_v24i7e39590_app2.png]

References

1. Galmiche S, Rahbe E, Fontanet A, Dinh A, Bénézit F, Lescure F, et al. Implementation of a self-triage web application for suspected COVID-19 and its impact on emergency call centers: observational study. J Med Internet Res 2020 Nov 23;22(11):e22924 [FREE Full text] [doi: 10.2196/22924] [Medline: 33147165]
2. Denis F, Galmiche S, Dinh A, Fontanet A, Scherpereel A, Benezit F, et al. Epidemiological observations on the association between anosmia and COVID-19 infection: analysis of data from a self-assessment web application. J Med Internet Res 2020 Jun 11;22(6):e19855 [FREE Full text] [doi: 10.2196/19855] [Medline: 32496206]
3. Denis F, Septans AL, Periers L, Maillard JM, Legoff F, Gurden H, et al. Olfactory training and visual stimulation assisted by a web application for patients with persistent olfactory dysfunction after SARS-CoV-2 infection: observational study. J Med Internet Res 2021 May 27;23(5):e29583 [FREE Full text] [doi: 10.2196/29583] [Medline: 34003765]
4. ORCHA. URL: https://orchahealth.com/ [accessed 2022-01-01]
5. Genève C. État des lieux de la E-santé en 2020, étude d'une application mobile de santé [thesis]. Sciences pharmaceutiques. 2020 Oct 19. URL: https://dumas.ccsd.cnrs.fr/dumas-02971517 [accessed 2022-06-16]
6. Di Maio M, Basch E, Denis F, Fallowfield L, Ganz P, Howell D, ESMO Guidelines Committee. The role of patient-reported outcome measures in the continuum of cancer clinical care: ESMO Clinical Practice Guideline. Ann Oncol 2022 Apr 18. [doi: 10.1016/j.annonc.2022.04.007] [Medline: 35462007]
7. Évaluation des applications dans le champ de la santé mobile (mHealth): État des lieux et critères de qualité du contenu médical pour le référencement des services numériques dans l'espace numérique de santé et le bouquet de services professionnels. Haute Autorité de Santé. 2021 Jun 30. URL: https://tinyurl.com/yarxcdwx [accessed 2022-06-16]
8. Riezebos RJ. Peer-reviewing of mHealth applications: requirements for peer-reviewing mobile health applications and development of an online peer review tool [thesis]. University of Amsterdam. 2014 Oct 01. URL: http://dare.uva.nl/cgi/arno/show.cgi?fid=573074 [accessed 2022-06-16]
9. Azad-Khaneghah P, Neubauer N, Miguel Cruz A, Liu L. Mobile health app usability and quality rating scales: a systematic review. Disabil Rehabil Assist Technol 2021 Oct;16(7):712-721. [doi: 10.1080/17483107.2019.1701103] [Medline: 31910687]
10. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, WHO mHealth Technical Evidence Review Group. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. BMJ 2016 Mar 17;352:i1174. [doi: 10.1136/bmj.i1174] [Medline: 26988021]
11. Digital Technology Assessment Criteria (DTAC). National Health Service. URL: https://www.nhsx.nhs.uk/key-tools-and-info/digital-technology-assessment-criteria-dtac/ [accessed 2022-06-16]
12. My Health Apps and blog. Patient View. URL: https://www.patient-view.com/my-health-apps/ [accessed 2022-01-01]
13. Medical Digital Solution Score. Institut National de la e-Santé. 2022 Mar 01. URL: http://med-clic.fr/ [accessed 2022-06-17]
14. Référentiel de bonnes pratiques sur les applications et les objets connectés en santé (mobile Health ou mHealth). Haute Autorité de Santé. 2016 Nov 07. URL: https://www.has-sante.fr/jcms/c_2681915/fr/referentiel-de-bonnes-pratiques-sur-les-applications-et-les-objets-connectes-en-sante-mobile-health-ou-mhealth [accessed 2022-06-16]
15. La version 2021 de la doctrine du numérique en santé est en ligne !. Agence du Numérique en Santé. 2022 Feb 04. URL: https://esante.gouv.fr/espace-presse/la-version-2021-de-la-doctrine-du-numerique-en-sante-est-en-ligne [accessed 2022-04-25]
16. Council Directive 93/42/EEC of 14 June 1993 concerning medical devices. European Commission. 1993 Jun 14. URL: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CONSLEG:1993L0042:20071011:en:PDF [accessed 2022-06-16]
17. Denis F, Krakowski I. How should oncologists choose an electronic patient-reported outcome system for remote monitoring of patients with cancer? J Med Internet Res 2021 Sep 09;23(9):e30549 [FREE Full text] [doi: 10.2196/30549] [Medline: 34499046]
18. Georges JL, Cheggour S. [Telemedicine, remote medicine, e-health:are French cardiologists ready?]. Ann Cardiol Angeiol (Paris) 2021 Nov;70(5):267-268. [doi: 10.1016/j.ancard.2021.09.015] [Medline: 34635332]
19. Denis F, Basch E, Septans A, Bennouna J, Urban T, Dueck AC, et al. Two-year survival comparing web-based symptom monitoring vs routine surveillance following treatment for lung cancer. JAMA 2019 Jan 22;321(3):306-307 [FREE Full text] [doi: 10.1001/jama.2018.18085] [Medline: 30667494]
20. Lin MH, Yuan WL, Huang TC, Zhang HF, Mai JT, Wang JF. Clinical effectiveness of telemedicine for chronic heart failure: a systematic review and meta-analysis. J Investig Med 2017 Jun;65(5):899-911 [FREE Full text] [doi: 10.1136/jim-2016-000199] [Medline: 28330835]
21. Niveau de preuve et gradation des recommandations de bonnes pratiques. Haute Autorité de Santé. 2013 Apr. URL: https://www.has-sante.fr/upload/docs/application/pdf/2013-06/etat_des_lieux_niveau_preuve_gradation.pdf [accessed 2022-06-16]
22. GGD Appstore. URL: https://www.ggdappstore.nl/ [accessed 2022-06-16]
23. Health Navigator. URL: https://www.healthnavigator.org.nz/apps/a/app-library/ [accessed 2022-06-16]
24. One Mind PsyberGuide. URL: https://onemindpsyberguide.org/apps/ [accessed 2022-06-16]
25. Osservatorio APP sanitarie. URL: http://www.appsanitarie.it/ [accessed 2022-06-16]
26. HealthOn. URL: https://www.healthon.de/ [accessed 2022-06-16]

27. 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]
28. Leigh S, Ouyang J, Mimnagh C. Effective? engaging? secure? applying the ORCHA-24 framework to evaluate apps for chronic insomnia disorder. Evid Based Ment Health 2017 Nov;20(4):e20. [doi: 10.1136/eb-2017-102751] [Medline: 28947676]
29. Baumel A, Faber K, Mathur N, Kane JM, Muench F. Enlight: a comprehensive quality and therapeutic potential evaluation tool for mobile and web-based eHealth interventions. J Med Internet Res 2017 Mar 21;19(3):e82 [FREE Full text] [doi: 10.2196/jmir.7270] [Medline: 28325712]
30. Henson P, David G, Albright K, Torous J. Deriving a practical framework for the evaluation of health apps. Lancet Digit Health 2019 Jun;1(2):e52-e54 [FREE Full text] [doi: 10.1016/S2589-7500(19)30013-5] [Medline: 33323229]
31. Institut National de la e-Santé. URL: https://ines-france.fr/ [accessed 2022-06-16]

Abbreviations

ESMO: European Society of Medical Oncology
HAS: Haute Autorité de Santé (French National Health Authority)
MDS: Medical Digital Solution

Edited by T Leung; submitted 16.05.22; peer-reviewed by M Elbaz, V Vercamer, ER Khalilian; comments to author 01.06.22; revised version received 08.06.22; accepted 14.06.22; published 05.07.22.

Please cite as:
Wagneur N, Callier P, Zeitoun JD, Silber D, Sabatier R, Denis F
Assessing a New Prescreening Score for the Simplified Evaluation of the Clinical Quality and Relevance of eHealth Apps: Instrument Validation Study
J Med Internet Res 2022;24(7):e39590
URL: https://www.jmir.org/2022/7/e39590
doi:10.2196/39590
PMID:35788102

Original Paper

The Effect of Tailored, Daily, Smartphone Feedback to Lifestyle Self-Monitoring on Weight Loss at 12 Months: the SMARTER Randomized Clinical Trial

Lora E Burke[1], MPH, PhD; Susan M Sereika[1], PhD; Zhadyra Bizhanova[2], MBChB, MPH; Bambang Parmanto[3], PhD; Jacob Kariuki[1], PhD; Jessica Cheng[2], MPH; Britney Beatrice[1], MS, RD; Maribel Cedillo[4], MS, RD; I Wayan Pulantara[3], PhD; Yuhan Wang[3], MSIS; India Loar[1], MPH; Molly B Conroy[4], MPH, MD

[1]School of Nursing, University of Pittsburgh, Pittsburgh, PA, United States

[2]School of Public Health, University of Pittsburgh, Pittsburgh, PA, United States

[3]School of Rehabilitation Sciences, University of Pittsburgh, Pittsburgh, PA, United States

[4]School of Medicine, University of Utah, Salt Lake City, UT, United States

Corresponding Author:
Lora E Burke, MPH, PhD
School of Nursing
University of Pittsburgh
3500 Victoria St
Pittsburgh, PA, 15261
United States
Phone: 1 412 779 0555
Email: lbu100@pitt.edu

Abstract

Background: Self-monitoring (SM) is the centerpiece of behavioral weight loss treatment, but the efficacy of smartphone-delivered SM feedback (FB) has not been tested in large, long-term, randomized trials.

Objective: The aim of this study was to establish the efficacy of providing remote FB to diet, physical activity (PA), and weight SM on improving weight loss outcomes when comparing the SM plus FB (SM+FB) condition to the SM-only condition in a 12-month randomized controlled trial. The study was a single-site, population-based trial that took place in southwestern Pennsylvania, USA, conducted between 2018 and 2021. Participants were smartphone users age ≥18 years, able to engage in moderate PA, with a mean BMI between 27 and 43 kg/m^2.

Methods: All participants received a 90-minute, one-to-one, in-person behavioral weight loss counseling session addressing behavioral strategies, establishing participants' dietary and PA goals, and instructing on use of the PA tracker (Fitbit Charge 2), smart scale, and diet SM app. Only SM+FB participants had access to an investigator-developed smartphone app that read SM data, in which an algorithm selected tailored messages sent to the smartphone up to 3 times daily. The SM-only participants did not receive any tailored FB based on SM data. The primary outcome was percent weight change from baseline to 12 months. Secondary outcomes included engagement with digital tools (eg, monthly percentage of FB messages opened and monthly percentage of days adherent to the calorie goal).

Results: Participants (N=502) were on average 45.0 (SD 14.4) years old with a mean BMI of 33.7 (SD 4.0) kg/m^2. The sample was 79.5% female (n=399/502) and 82.5% White (n=414/502). At 12 months, retention was 78.5% (n=394/502) and similar by group (SM+FB: 202/251, 80.5%; SM: 192/251, 76.5%; *P*=.28). There was significant percent weight loss from baseline in both groups (SM+FB: –2.12%, 95% CI –3.04% to –1.21%, *P*<.001; SM: –2.39%, 95% CI –3.32% to –1.47%; *P*<.001), but no difference between the groups (–0.27%; 95% CI –1.57% to 1.03%; *t* =–0.41; *P*=.68). Similarly, 26.3% (66/251) of the SM+FB group and 29.1% (73/251) of the SM group achieved ≥5% weight loss (chi-square value=0.49; *P*=.49). A 1% increase in FB messages opened was associated with a 0.10 greater percent weight loss at 12 months (b=–0.10; 95% CI –0.13 to –0.07; *t* =–5.90; *P*<.001). A 1% increase in FB messages opened was associated with 0.12 greater percentage of days adherent to the calorie goal per month (b=0.12; 95% CI 0.07-0.17; *F*=22.19; *P*<.001).

Conclusions: There were no significant between-group differences in weight loss; however, the findings suggested that the use of commercially available digital SM tools with or without FB resulted in a clinically significant weight loss in over 25% of participants. Future studies need to test additional strategies that will promote greater engagement with digital tools.

Trial Registration: Clinicaltrials.gov NCT03367936; https://clinicaltrials.gov/ct2/show/NCT03367936

(J Med Internet Res 2022;24(7):e38243) doi:10.2196/38243

KEYWORDS

self-monitoring; behavioral intervention/weight loss; feedback messages; engagement; mHealth; adherence; obesity; randomized clinical trial; smart scales; physical activity trackers; digital health

Introduction

Obesity is associated with several chronic diseases [1,2]. Obesity prevalence in the United States exceeds 42.4% and disproportionately affects racial and ethnic minority groups [3,4].

The gold standard for weight loss treatment is standard behavioral treatment (SBT), which includes reduced energy intake, increased energy expenditure, and in-person, group-based behavioral counseling plus feedback (FB) on self-monitoring (SM) from a trained interventionist [5-7]. However, SBT is difficult to implement on a large scale to reach populations most in need of treatment [8]. There is a critical need for more affordable, scalable, less burdensome, and efficacious treatments for weight loss [4].

The cornerstone of SBT is SM with interventionist FB [9-12]. A meta-regression demonstrated that SM use was the strongest predictor of efficacy in a weight loss intervention [13]. The highest efficacy was reported in a study in which SM was combined with another self-regulation technique [14], such as FB. Several studies have examined strategies to enhance sustained engagement in SM, including use of digital tools [5,11,12,15-20].

However, despite improvements in SM (eg, advancing from paper to digital tools), 2 issues persist: individuals still find SM burdensome [21] and SM adherence declines over time, which is associated with poorer weight loss outcomes [11,15,17-19,21-28]. Advances in mobile technology provide opportunities to enhance interventions targeting SM, expand their reach, and prevent decline in adherence. Delivering real-time FB to SM can reinforce behavior change [29] and partially replace in-person sessions [30,31]. The addition of wearable activity trackers [32] and smart scales [33] that synchronize data with a smartphone eliminates the need to manually record physical activity (PA) and weight, reducing burden and increasing adherence [19,34,35].

We previously examined the effect of providing FB to dietary SM and PA; however, the hardware and software used were rudimentary compared to today's technology [16]. Despite these limitations, remotely delivered FB messages enhanced SM adherence and improved weight loss [29,36]. These results and significant mobile technology enhancements provided groundwork for the expanded algorithm and FB intervention used in this trial, SMARTER [23].

SMARTER was a 2-group randomized controlled trial that enrolled 502 adults with random assignment to either (1) SM alone (n=251) or (2) SM+FB (n=251) and examined the efficacy of the approaches. We examined short-term outcomes at 6 months with 86% retention that demonstrated that both groups lost a significant percent weight from baseline (SM+FB: –3.16%, $P<.001$; SM: –3.20%, $P<.001$) but found no significant between-group difference ($P=.94$) [37]. We hypothesized that the SM+FB group would show greater percent weight loss at 12 months compared to the SM group. For the SM+FB group, we also report the percent of FB messages opened and association with weight change and daily calorie goal adherence. We hypothesized that a greater number of FB messages that were opened would result in greater weight loss and adherence to the calorie goal.

Methods

Ethical Considerations

We published the study protocol and design previously [23]. The study was approved by the Institutional Review Board for Human Protection at the University of Pittsburgh (#19060112) and registered on ClinicalTrials.gov (NCT03367936). We informed all participants of screening procedures prior to obtaining consent and obtained in-person informed consent for the intervention study.

Recruitment

Recruitment, conducted in the greater community surrounding Pittsburgh, PA, commenced in August 2018 and ended in March 2020. The intervention trial was completed in April 2021. Both online and in-person methods were used. Initially, interested individuals who were regular smartphone users completed surveys and a 5-day food diary in which they needed to record at least 700 calories of food intake per day to ensure that they could SM. Once deemed eligible, individuals had an in-person assessment to verify weight and height for BMI measures. Inclusion criteria were BMI between 27 and 43 kg/m^2, completion of a 5-day electronic food diary, and ability to engage in moderate PA. Exclusion criteria were needing supervision of diet or PA, pregnancy, serious mental illness (eg, schizophrenia), alcohol abuse or eating disorder, and current weight loss treatment [23].

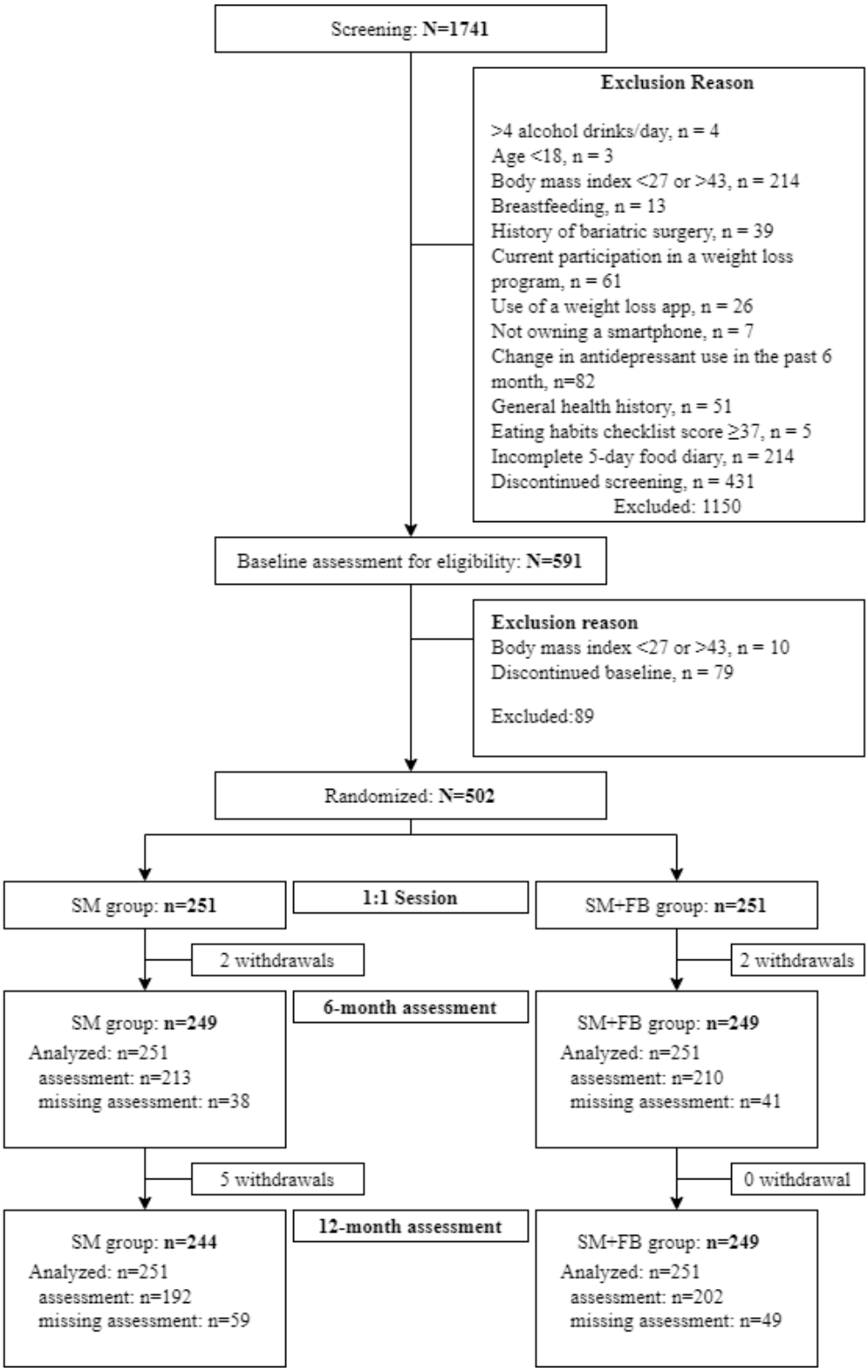
Randomization

After completing the intervention consent, research staff used a randomization software program to determine group assignment that was generated using minimization with

stratification by gender (male or female) and race (Black or non-Black) with equal allocation to the 2 treatment conditions (please see the CONSORT [Consolidated Standards of Reporting Trials] diagram in Figure 1). All randomized participants were included in the final analyses. We used Wadden's conservative approach for imputing missing weight data assuming that there was a weight gain.

Key staff (BB, IL) who conducted the assessments were not blinded to the treatment assignment, whereas all other personnel and investigators, including the statisticians, (LEB, SMS, ZB, BP, JK, JC, MC, IWP, YW, and MBC) were blinded to assignment. Since participants were informed of both treatment conditions, they could not be blinded.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram for the SMARTER trial.



Intervention

Behavioral Intervention

The intervention was grounded in behavioral change theory with an emphasis on Kanfer's self-regulation theory that posits that SM is central to behavior change and includes FB tailored to the SM data. At baseline, all participants had a 90-minute, one-on-one, in-person intervention session with a dietitian on the core concepts of SBT followed by a demonstration of the Fitbit app to enter foods eaten for SM of diet, a Fitbit activity tracker to monitor PA, and a smart scale for daily self-weighing. Use of the investigator-developed SMARTER app, which was used only for random retrieval of FB messages from the message library and delivery of messages to the participant's smartphone, was demonstrated to the SM+FB participants, so they could view the prompt icon for the FB messages and open the app to read the message. Participants used their own smartphones; the other SM devices (Fitbit activity tracker and commercially available smart scale) were provided by the study.

Dietary Intake

Participants used the Fitbit app to view food nutrient values, app-generated subtotals, and the daily intake summaries. The calorie goal was determined from baseline body weight (women: 1200 kcal for <200 lb or 1500 kcal ≥200 lb; men: 1500 kcal for <200 lb or 1800 kcal for ≥200 lb) and individualized as needed [23]. Fat gram goals approximated 25% of the calorie goal (eg, 33 or 42 grams per day for females).

Physical Activity

All participants monitored PA using a wrist-worn activity tracker, the Fitbit Charge 2, synced with their smartphone. Staff instructed participants to increase their PA gradually, primarily by walking, and to aim for 150 minutes per week by 12 weeks [38]. Once at goal, they were encouraged to add 10 minutes per week until they reached 300 minutes per week [23]. All aerobic activities counted toward PA goals. The Fitbit database stored total steps, sedentary minutes, and active minutes.

Weight

All participants were instructed to weigh daily on the study-provided smart scale that transmitted weight data to their smartphone and study database.

Feedback Messages

The FB algorithm was programmed on the study's server and used real-time synced SM data to send the FB message up to 3 times per day. Individuals in the SM+FB group received up to 3 FB messages per day on their smartphone during waking hours tailored to the most recent SM data and addressing caloric as well as fat and added-sugar intake daily and PA every other day. Weekly weight FB was based on whether self-weighing occurred and the amount or rate of weight change. FB messages addressed 1 behavior at a time. The participant received a prompt that there was a new FB message on the smartphone. If the FB message was not opened within 1 hour of being sent, the SMARTER icon prompt and message disappeared; if the message was opened, the participant could save it for future review. More details on the FB messages and study infrastructure for message delivery are available elsewhere [23,37].

Engagement with SM tools was a crucial component of the intervention, as the algorithm used the SM data to determine an appropriate FB message. If the participant did not SM, FB messages were sent encouraging SM. After 2 weeks of missing SM data, staff sent an email query about technical issues and encouraged SM. Additional details on the algorithm and FB messages are published elsewhere [23]. The message library was changed at least monthly to avoid participant desensitization to FB [29]. Individuals in the SM group did not receive FB messages or staff emails. Further details of all intervention components and the algorithm driving the FB messages have been published elsewhere [23].

Outcomes

Assessments

Both in-person and web-based assessments were used initially, and the in-person assessments were performed at the Clinical Research Center in the School of Nursing at the University of Pittsburgh. At in-person assessments, participants (in light clothing and bare feet) stood on a Tanita scale and body fat analyzer at baseline, 6 months, and 12 months. Percent lean and fat mass were also collected; however, after March 16, 2020 (ie, the beginning of the COVID-19 pandemic shutdown), we collected 12-month weight data remotely from the participants' study-provided scales, which assessed only weight and percent fat mass. Staff contacted participants to instruct them to dress in clothing as in the baseline assessment and report their weight, which was also captured electronically [23]. At 12 months, of the 502 participants, 189 (37.7%) had in-person weights, 205 (40.8%) had remote weights, and 108 (21.5%) had missing weights. Smart scale weights recorded within 2 weeks of 6- or 12-month assessments were used for imputation of missing weights. If no weight was recorded by the smart scale, a 0.01 kg/day weight gain was assumed from last available scale weight value [39]. Cardiometabolic measures are reported in Multimedia Appendix 1 (Figures S1-S6). Participants were compensated for completing the 6- and 12-month assessments regardless of mode of conduct, in-person or remotely.

Feedback Messages

The monthly percentage of FB messages opened over the 12 months are expressed as the number of FB messages opened over the total number of FB messages sent in 30-day increments and multiplied by 100%.

Adherence to the Study-Defined Calorie Goal

The monthly percentage of days adherent to the calorie goal (defined as between 85% and 115% of daily calorie goal) among participants meeting the dietary SM goal (≥50% daily calorie goal) was calculated as follows:

> Number of days meeting calorie goal per month / Number of days meeting dietary SM criteria per month × 100%

Statistical Analysis

The planned total sample size for this randomized controlled trial was determined as 530 (265 per treatment group), allowing for a statistical power of 0.80 to detect effect sizes (standardized mean differences, *d*) as small as *d*=0.301 for the mean percent weight changes at 6 and 12 months between the SM-only and SM + FB groups when using linear mixed modeling with linear contrasts at a Bonferroni-adjusted significance level of *P*=.025 and for at most 20% attrition [23]. Due to the COVID-19 pandemic, recruitment was stopped in March 2020 with 502 randomized participants (251 per treatment arm). With this reduced sample size, slightly larger small-to-medium effect sizes of *d*=0.309 would still be detectable with 0.80 power at an adjusted significance level of *P*=.025, allowing for up to 20% attrition.

Continuous variables are summarized as mean and SD, and descriptive statistics for categorical variables are reported as counts with percentages. Appropriate group comparative analyses were performed on participant descriptors and outcome variables at baseline by randomized treatment assignment [37]. The effect of treatment assignment on percent weight change over 12 months was examined using linear mixed modeling following intention to treat. Models included a random intercept and unstructured variance-covariance matrix for the repeated assessments, supported by Akaike information and Bayesian information criteria. The base model included fixed effects for time (baseline vs 6 months and 12 months), group (SM+FB vs SM alone), and group by time interaction. Sensitivity analyses were performed among completers only and using inverse probability weighting for dropout by 12 months with age, race, and baseline as predictors.

The effect of the percentage of FB messages opened on percent weight change from baseline to 12 months for the SM+FB group was analyzed using univariate linear regression. Additionally, the associations of monthly percentages of days adherent to the calorie goal with treatment assignment and monthly percentages of FB messages opened were analyzed using separate linear mixed models with random intercept and slope for the total sample and for the SM+FB group, respectively. We conducted sensitivity analyses on the treatment effects on monthly percentages of days adherent to the calorie goal over 12 months in the total sample and on the associations between monthly percentages of FB messages opened and monthly percentages of days adherent to the calorie goal in the SM+FB group for the varying monthly percentage of days with sufficient dietary SM data (data not shown). Here, we report the results using days with ≥50% of the calorie goal recorded or ≥15 of 30 days with sufficient dietary SM data.

Model assessment (ie, residual analyses with influence diagnostics) was performed for each fitted model; sensitivity analyses were conducted for outlying or influential observations and to explore the effect of the COVID-19 pandemic on the efficacy of treatment assignment on percent weight change (data not shown). All analyses were performed using SAS version 9.4 (SAS Institute).

Results

Baseline Characteristics

Most participants were White (414/502, 82.5%), female (399/502, 79.5%) and on average 45.0 (SD 14.4) years old. Sociodemographic, clinical, and psychosocial characteristics, as well as primary outcome measurements at baseline, were similar between the treatment groups [37].

Retention

At 12 months, the overall retention was 78.5% (394/502). Retention was similar by treatment condition, with SM+FB at 80.5% (202/251) and SM at 76.5% (192/251; X^2=1.18; *P*=.28).

Percent and Absolute Weight Change

Figures 2 and 3 illustrate the results from linear mixed modeling for the effect of treatment assignment on weight loss and percentage of weight loss over 12 months. On average, both groups had statistically significant weight loss over 12 months ($b_{6\ months}$=–2.94, 95% CI –3.70 to –2.19; $b_{12\ months}$=–2.34, 95% CI –3.10 to –1.59; *F*= 61.46; *P*<.001). The trajectory of weight change over time was similar between groups ($b_{group \times 6\ months}$=0.09, 95% CI –0.97 to –1.16; $b_{group \times 12\ months}$=0.36, 95% CI –0.70 to –1.43; *F*=0.24; *P*=.79), and there were no significant overall treatment effects on weight change (b_{group}=–0.32; 95% CI –3.04 to 2.40; *F*=0.02; *P*=.*90*). There was a significant percent weight loss from baseline in both groups (SM+FB: –2.12%, 95% CI –3.04% to –1.21%, *P*<.001; SM: –2.39%, 95% CI –3.32% to –1.47%, *P*<.001), but no difference between the groups (–0.27%; 95% CI –1.57% to 1.03%; *t*=–0.41; *P*=.68). The percentages of participants who lost ≥5% weight from baseline to 12 months were similar between the SM+FB (66/251, 26.3%) and SM (73/251, 29.1%) arms (X^2=0.49; *P*=.49). Based on sensitivity analyses using inverse probability weighting, there was no significant difference in percent weight change at 12 months in the SM+FB arm (mean –3.57%, SD 20.16) and the SM arm (mean –3.53%, SD 19.94; *t*=0.07; *P*=.95). Additional analyses among completers showed no significant treatment effects on percent weight change at 12 months (SM+FB: mean –3.54%, SD 7.16; SM mean: –3.58%, SD 7.06; *t*= –0.07; *P*=.95). For the total sample, mean weight change in kilograms at 12 months was mean –2.16 kg (SD 0.27). There was no significant difference in weight change in kilograms at 12 months in the SM+FB arm (mean –1.98 kg, SD 0.38) or the SM arm (mean –2.34, SD 0.38; 2-sample *t* test=0.67; *P*=.50).

Analyses without outliers had results similar to the full-sample results. There were no significant effects for the COVID-19 pandemic on the relationship of treatment with weight change over time.

Figure 2. Effect of treatment assignment on weight change.

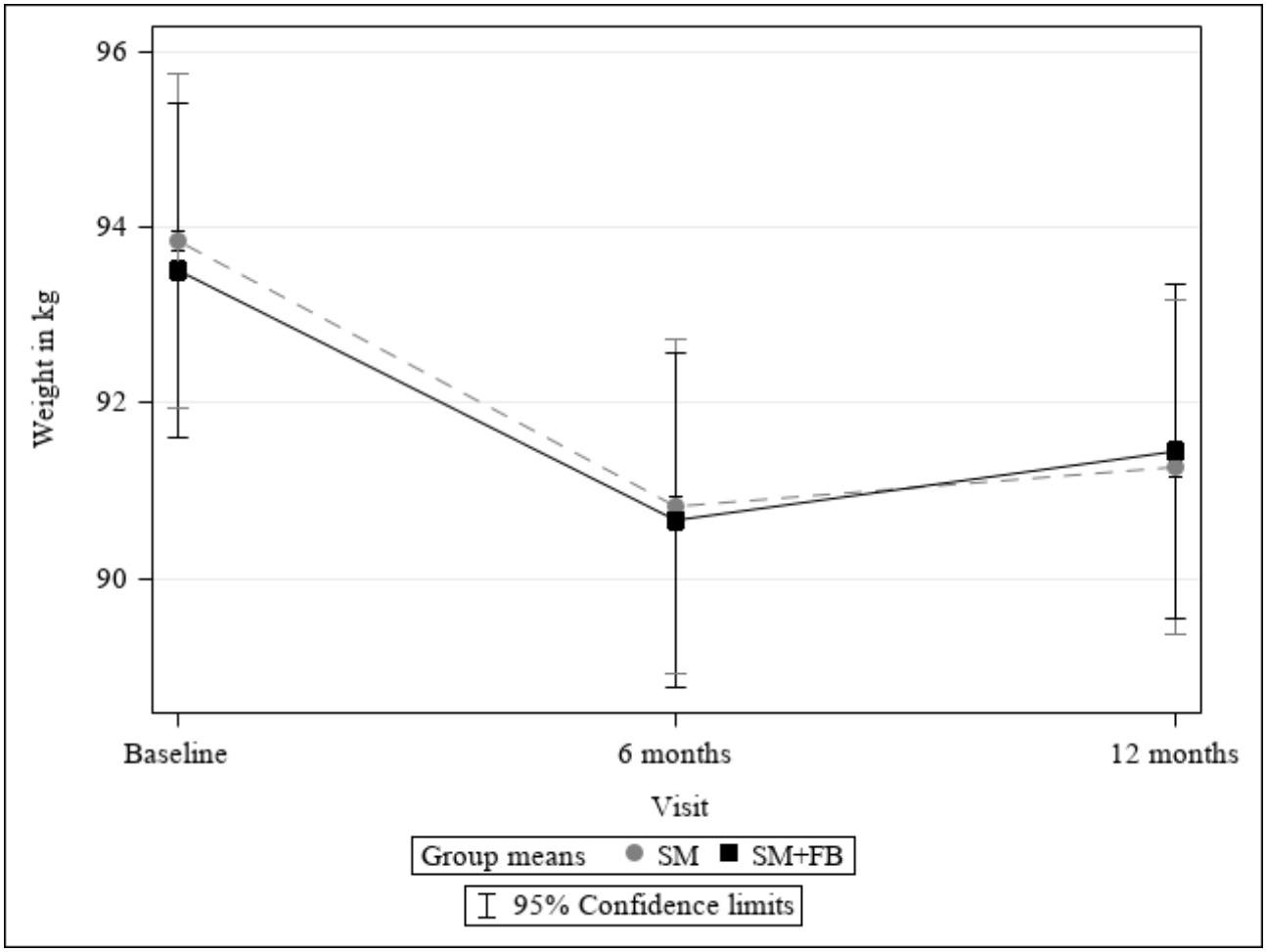
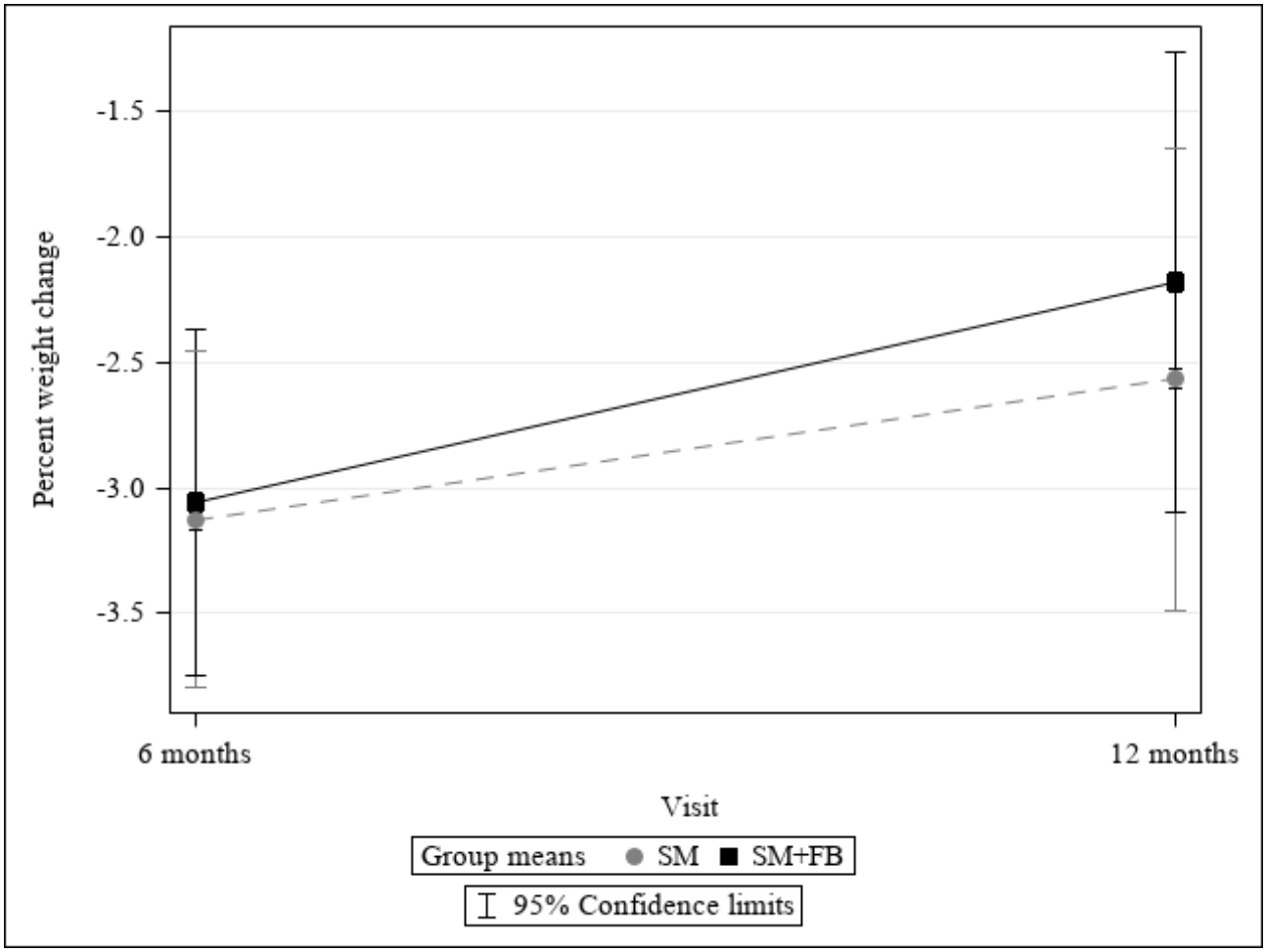


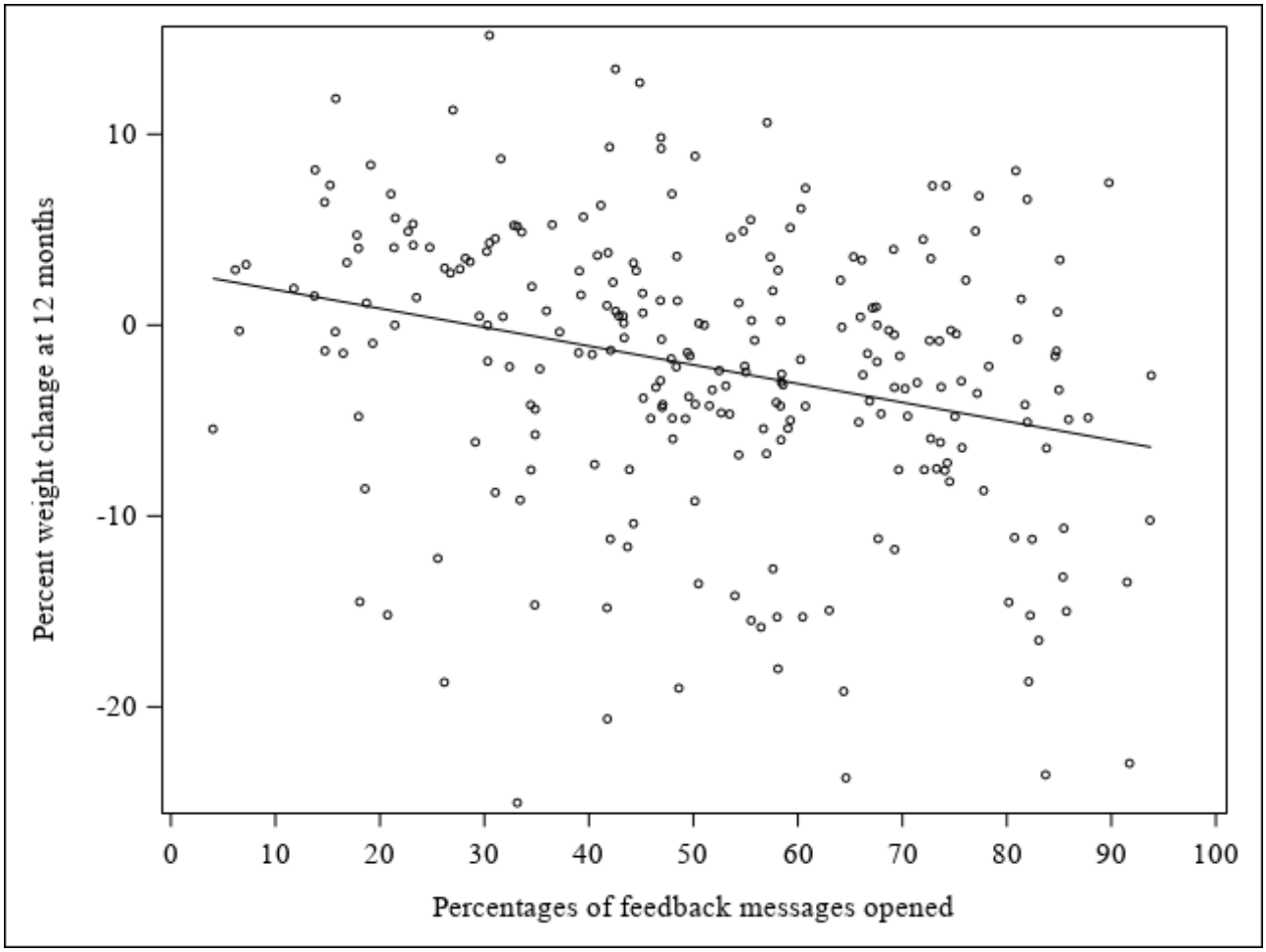
Figure 3. Effect of treatment assignment on percent weight change.



Feedback Messages Opened

In the SM+FB arm, the median percentage of FB messages opened from baseline to 12 months was 42.19% (461/1026; q1: 234/597, 39.20%; q3: 728/1095, 66.48%) and ranged from 1.28% (14/78) to 93.70% (1026/1095). Figure 4 displays the association of the percentage of FB messages opened and percent weight change at 12 months. A 1% increase in FB messages opened was associated with a 0.10 greater percent weight loss at 12 months (b=–0.10; 95% CI –0.13 to –0.07; t=–5.90; P<.001).

Figure 4. The associations between percentages of feedback messages opened and percent weight change at 12 months.



Percentage of Days Adherent to the Calorie Goal

Figure 5 illustrates the change in monthly percentage of days adherent to the calorie goal by treatment group over 12 months using the criteria of recording ≥50% of the calorie goal 15 days per month. The monthly percentage of days adherent to the calorie goal declined nonlinearly in both groups. Overall, the percentage of days adherent to the calorie goal was greater in the SM+FB group than in the SM group (b_{group}=4.43; 95% CI 0.41-8.45; *F*=4.67; *P*=.03). The rate of decline in percentage of days adherent to the calorie goal was slower in the SM+FB arm than the SM arm over 12 months ($b_{group\times\ time\text{-}linear}$=–1.98, 95% CI –3.03 to –0.93, *F*=13.71, *P*<.001; $b_{group\times\ time\text{-}quadratic}$=0.14, 95% CI 0.06-0.22, *F*=11.04, *P*<.001).

Figure 5. Effect of treatment assignment on percentage of days adherent to the calorie goal over 12 months.

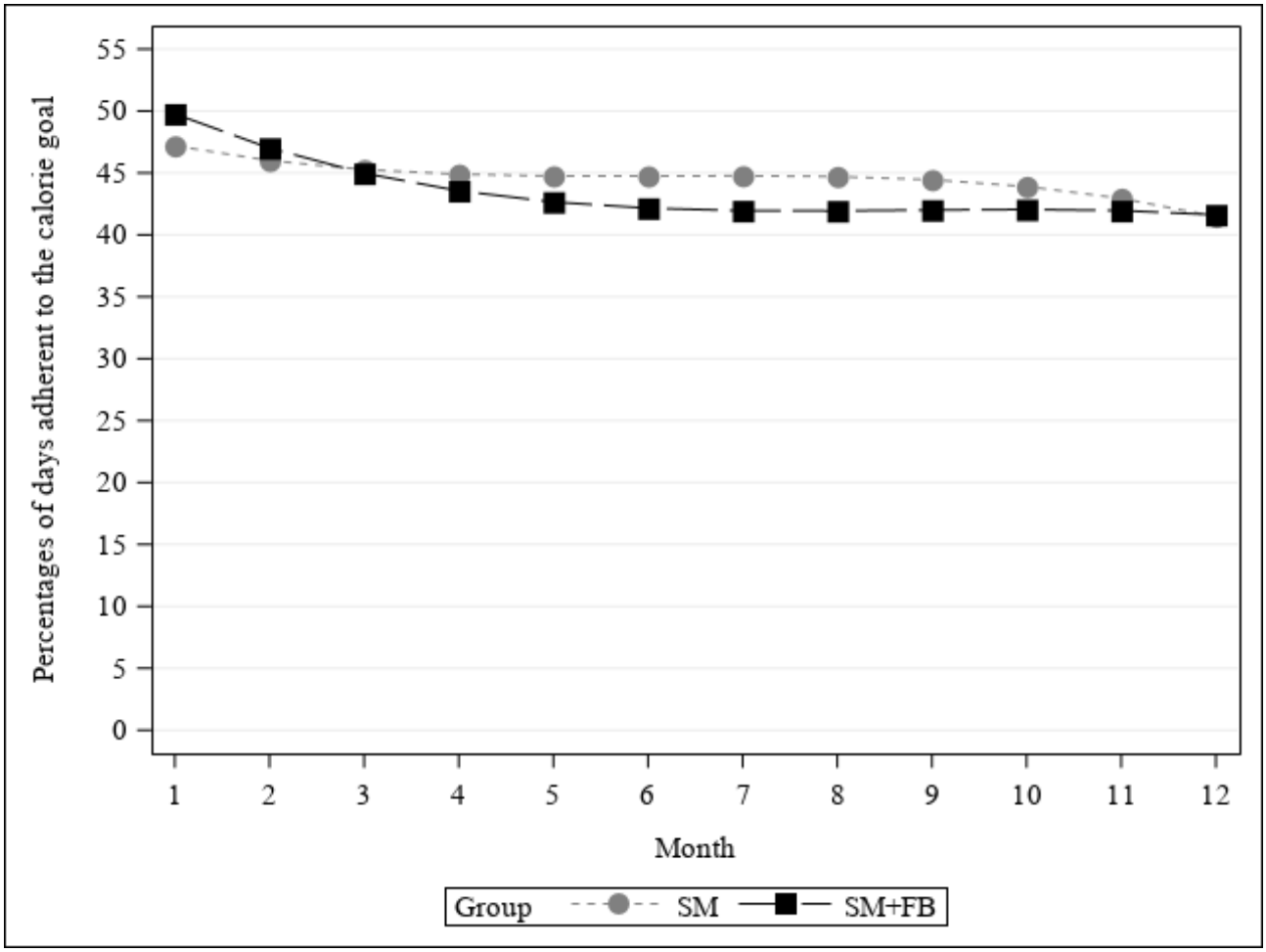


Figure 6 illustrates the change in monthly percentages of FB messages opened and monthly percentages of days adherent to the calorie goal over 12 months and their association in the SM+FB group, respectively. In general, monthly percentages of FB messages opened ($b_{time\text{-}linear}$=–8.34, 95% CI –9.91 to –6.78, F=110.26, P<.001; $b_{time\text{-}quadratic}$=0.54, 95% CI 0.27-0.81, F=15.07, P<.001; $b_{time\text{-}cubic}$=–0.02, 95% CI –0.04 to –0.009, F=10.63, P=.001) and monthly percentages of days adherent to the calorie goal declined nonlinearly over 12 months ($b_{time\text{-}linear}$=–3.37, 95% CI –5.29 to –1.45, F=11.93, P=.001; $b_{time\text{-}quadratic}$=0.44, 95% CI 0.10-0.79, F=6.26, P=.01; $b_{time\text{-}cubic}$=–0.02, 95% CI –0.04 to 0.001, F=3.62, P=.06), with a greater percentage of FB messages opened being associated with higher adherence to the calorie goal as shown in Figure 7 (b_{FB}=0.12; 95% CI 0.07 to 0.17; F=22.19; P<.001). There was no significant interaction between the percentage of FB messages opened and the polynomial time effects.

Figure 6. Change in monthly percentages of feedback messages opened and days adherent to the calorie goal in SM+FB over 12 months.

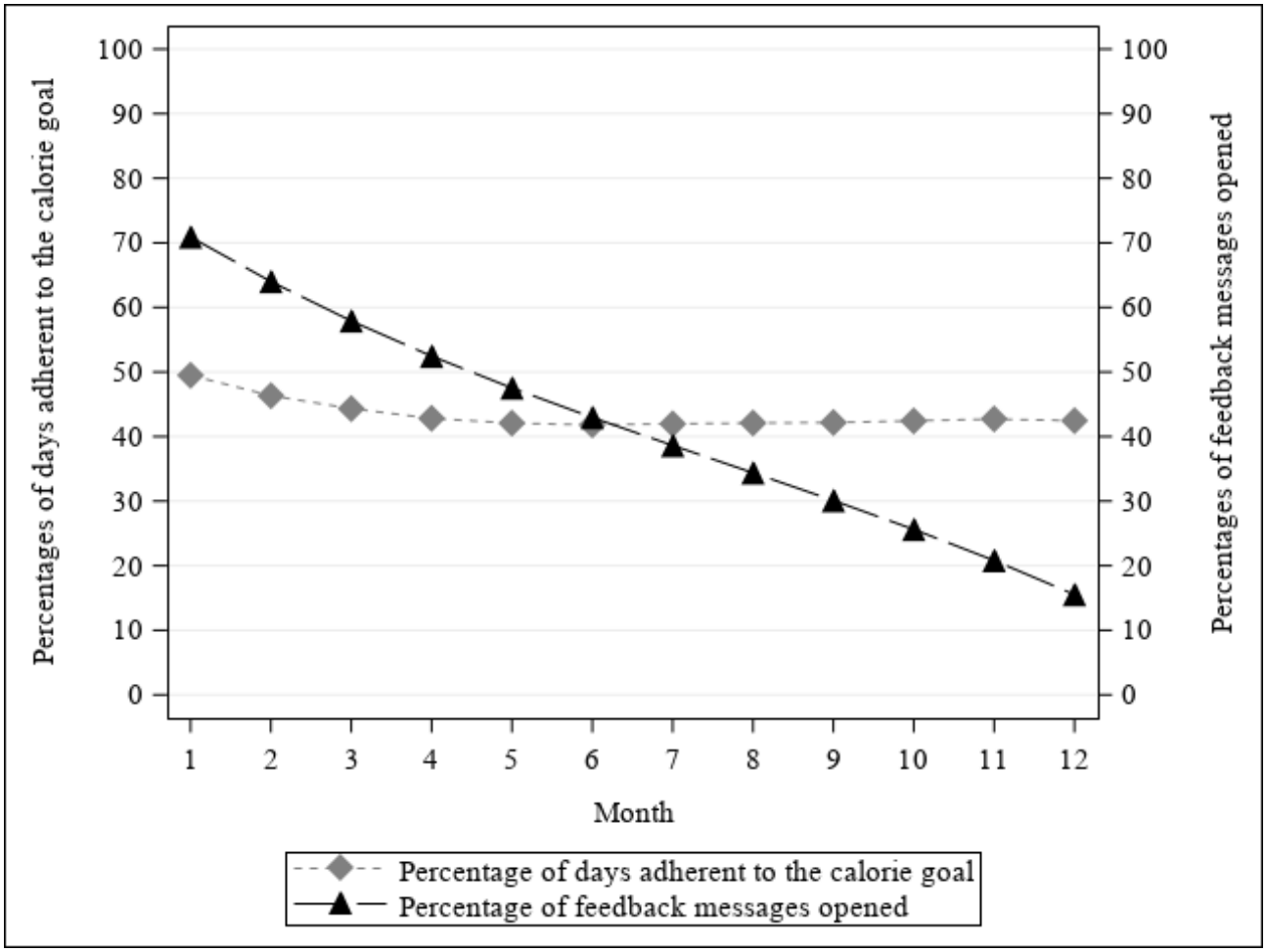
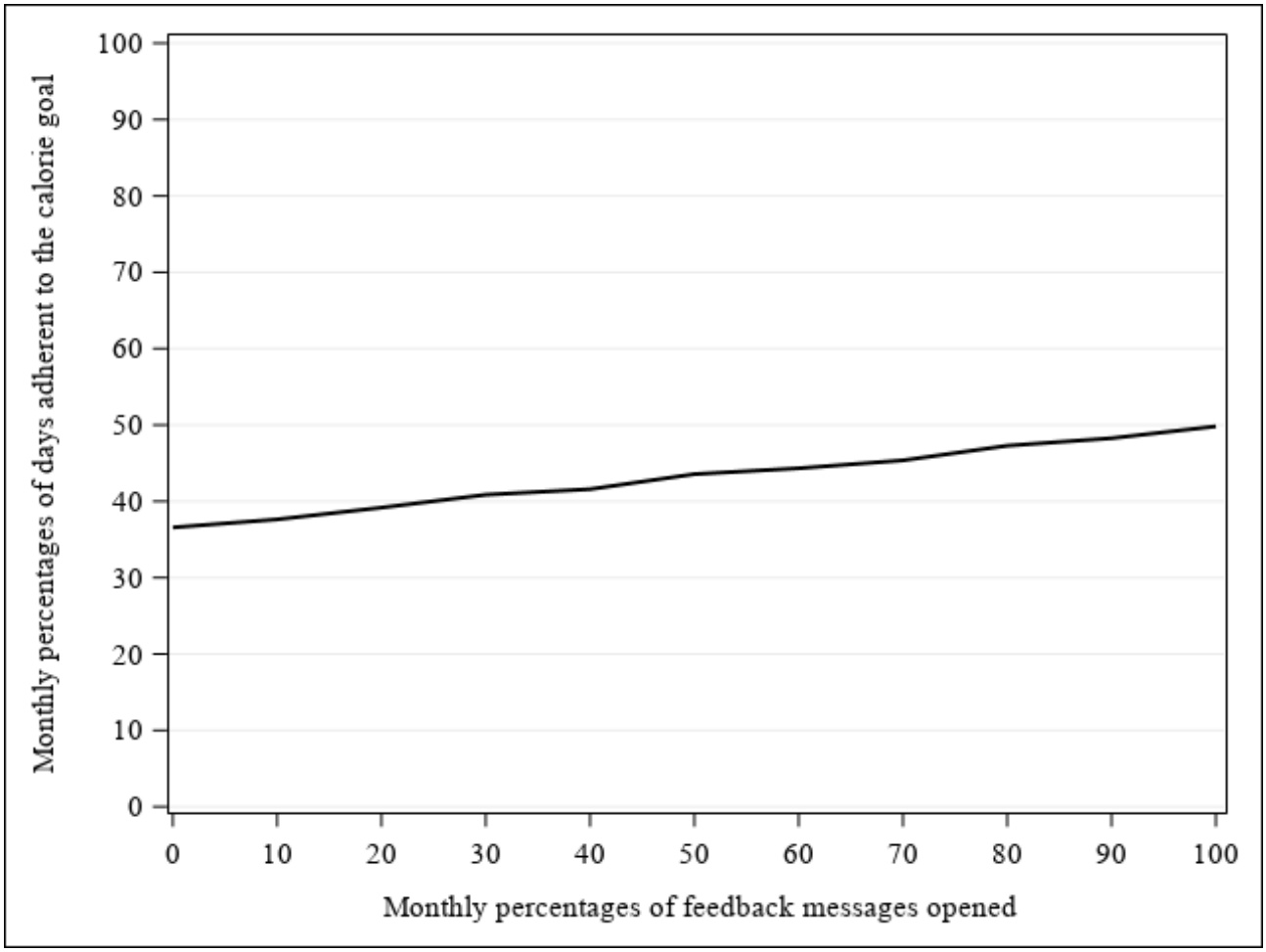


Figure 7. The associations between monthly percentages of feedback messages opened and days adherent to the calorie goal in SM+FB over 12 months.



Discussion

Principal Results

We conducted a trial of a scalable, remotely delivered, behavioral weight loss intervention and tested the efficacy of a custom-developed, theoretically based, smartphone app that provided real-time FB remotely to reinforce diet, PA, and self-weighing behaviors. We observed a small but significant percent weight change from baseline to 12 months with no significant difference between the groups, suggesting the FB provided no additional benefit beyond SM. Our findings suggest that it is feasible to deliver a 12-month, remotely delivered intervention for weight loss to a large sample even with COVID-19 pandemic restrictions. The percent weight losses observed (equivalent to an absolute weight change of 2.0-2.5 kg) were less than what is usually reported in studies of in-person group counseling or online coaching [40-42] but not different from other small trials that tested digital interventions with limited human interaction [43]. A systematic review of eHealth interventions reported similar weight losses (1.4 to 2.7 kg) [44] at postintervention [45].

In this paper, we report initial app engagement findings, specifically the number of times participants opened the app to SM or read FB messages. Our study demonstrated that without personal interaction, engagement declined to the point that it interfered with the intervention delivery. Although the FB message system worked in theory, the declining adherence to SM created a lack of data for the algorithm to select a FB message. Sending an email message after 2 weeks without evident SM might have been too long to wait to prompt re-engagement, and a phone call might have been more effective; however, we were trying to increase the study scalability. The COVID-19 shutdown ended all in-person interactions including assessments, which might have affected engagement.

We measured engagement by the percentage of FB messages opened, which was reflective of dietary SM adherence. Overall, approximately 40% (461/1026, 42.19%; IQR 45.30) of the FB messages were opened with a very wide range of 1.28% (14/78) to 93.70% (14/78) in our study. Although this is less than ideal, engagement was equal (or greater) than the 41% to 60% rate of engagement that others have reported [46-48] in mobile health studies. A recent literature review emphasized the challenge of initial and sustained engagement in mobile health studies [49]. Considering the completely remote intervention with minimal to no in-person contact, our findings are encouraging but indicate a continued need to improve sustained engagement in SM [50].

Comparison With Prior Work

There was greater adherence to the calorie goal in the SM+FB group compared to the SM group, and a greater number of FB messages opened was associated with a greater percent weight

loss in the SM+FB group, suggesting that when messages were opened, the FB messages reinforced behavior changes related to food selection. These findings are similar to those in the SMART trial which showed that the personal data assistant + FB group had better adherence to the dietary goals and was the only 1 of the 3 study groups that had a significant within-group weight loss over 24-months. In SMART, adherence to SM strongly predicted weight loss at all time points [17,29,49].

The study algorithm and FB message library for SMARTER were significantly expanded from the earlier SMART trial; however, these improvements did not compensate for the absence of the 16 in-person group sessions that were part of the SMART trial, suggesting that some form of interpersonal interaction may be needed to augment mobile health interventions. Recent studies have reported similar findings [40,42]. For example, Thomas et al [40] demonstrated that providing a monthly in-person weigh-in to accompany 5-minute skills training videos achieved weight losses comparable to the gold standard of frequent in-person group sessions over the 18-month trial. Similarly, Amagai et al's [49] literature review suggested that coaching to provide social support is an important strategy to improve engagement.

The comparison group in SMARTER received the same treatment components (ie, a one-to-one, in-person intervention session at baseline and digital tools for SM) and achieved very similar weight loss without receiving any prompts or reminders to SM. The intent of this comparative intervention was to determine the effectiveness of the approach that thousands of individuals are using by purchasing apps and tracking devices. These results suggest that some individuals who receive individual guidance at baseline with goals for diet and PA and encouragement to SM can achieve a clinically significant weight loss under their own direction.

Both groups used the Fitbit Charge 2 for tracking diet and PA. The Fitbit provided graphical FB on dietary intake (total calories consumed and "burned") and a weekly summary of PA; thus, the SM-only group received some automatic FB if they were syncing their device to their phone. Although this FB could be reinforcing for some individuals, it lacked the personalized component that the SMARTER FB provided since the SMARTER messages were tailored to SM data entered at that time, were positive in tone, and often provided suggestions. However, due to the lower-than-expected engagement with SM, many individuals did not open (and therefore receive) enough messages.

A recent pilot study that used a 2 × 2 factorial design provides some insights into FB (counselor-crafted vs pre-scripted [51]) and group sessions (yes or no). Participants in the group sessions were more engaged in SM and lost more weight than did those not offered group sessions; however, the group that received pre-scripted, modular FB had significantly greater weight loss than did the group that received the counselor-crafted FB while there was no consistent difference in their treatment engagement. It is not known why the group that received briefer FB lost more weight; the longer FB sent weekly was possibly perceived as burdensome. The authors and other researchers suggested that the mechanisms underlying FB are poorly defined and that the amount, timing, frequency, and framing are just a few of the dimensions that need to be further studied [52-54].

Several recent weight loss intervention studies have examined an array of digital strategies to enhance adherence to SM while reducing components of the gold standard SBT; however, several had small samples, conducted brief interventions, and had small weight losses [26,43,55]. Despite these limitations, results showed promise for further study of approaches to enhance SM adherence (eg, counseling phone calls [26] or weekly emails with structured lessons [43]). The cumulative evidence makes it difficult to determine which intervention components can be most effective in producing clinically meaningful weight loss. Specifically, it is difficult to ascertain how much of the human interventionist component can be replaced to make weight loss treatments scalable to a broader reach and lower operational costs. This critical gap in the evidence needs to be addressed in future studies, so we can broaden our reach to the millions who need weight loss treatment, particularly those who do not have access to existing clinical and commercial weight loss programs.

Study Strengths

There are several strengths to our study: a large sample size, a rigorous randomized design with a comparable control group, a retention rate higher than that reported by shorter and similar studies [56], use of validated measures, defined adherence metrics, and an objective measure of FB messages opened. The theory-based intervention was expanded from a previously tested and efficacious FB system. Additional strengths include using an extensive remote screening system and, born of necessity due to the COVID pandemic, pivoting to a remotely delivered, objective assessment protocol with minimal data loss.

Study Limitations

Limitations include recruitment of fewer males and minorities than targeted, which limits generalizability. The retention was slightly lower than the targeted 80%. The metric of FB messages opened does not necessarily equate to the actual number of FB messages read.

Conclusions

The SMARTER trial delivered customized, real-time FB to participants based on SM data and capitalized on the use of available digital technology to provide personal weight management support without ongoing human counseling. This approach is scalable, as it reduces cost and participant burden while increasing reach to those without access to SBT or who do not wish to participate in an in-person program. We hypothesized that participants in the SM+FB group would have greater weight loss than those in the SM group at 12 months; however, weight loss outcomes were similar. Our results suggest that the addition of FB to SM did not make a significant between-group difference in weight loss outcomes; however, those who remained engaged and opened more FB messages had better calorie goal adherence and weight outcomes. Moreover, one-fourth and almost one-third of each group achieved clinically significant weight loss, suggesting that for a portion of participants, the SM component of the intervention was efficacious.

Considering the unrelenting prevalence of obesity, there is a critical need for scalable interventions that can reach those most at risk and with the least resources. The evidence supports standalone, scalable digital interventions, yet the crucial challenge is the development of digital tools that will keep users engaged long enough to see positive, sustainable outcomes. In advancing the digital aspects, we also need to identify the most efficacious personal interaction components that best augment and support sustained SM and lifestyle change. Obesity is a complex, multifactorial, chronic condition that requires ongoing support and an array of treatment options that will accommodate for the diverse needs of those seeking treatment.

Acknowledgments

This research was supported by grants from the National Institutes of Health (NIH) to LEB (#R01-HL131583), JK (#R01-HL131583S), and JC (#F31-HL15627), and a grant from the University of Pittsburgh Clinical Translational Science Institute to Dr Steve Reis (#UL1-TR001857).

The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or the National Heart, Lung, and Blood Institute.

The authors would also like to acknowledge the participants who enrolled in the study and the time they gave to our research.

Authors' Contributions

LEB wrote the manuscript, SMS and ZB conducted the data analyses, and MBC contributed guidance and consultation throughout the writing process. All authors read and approved the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Cardiometabolic measures.
[DOCX File , 191 KB - jmir_v24i7e38243_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.2).
[PDF File (Adobe PDF File), 108 KB - jmir_v24i7e38243_app2.pdf]

References

1. Jensen MD, Ryan DH, Apovian CM, Ard JD, Comuzzie AG, Donato KA, et al. 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]
2. LeBlanc ES, Patnode CD, Webber EM, Redmond N, Rushkin M, O'Connor EA. Behavioral and pharmacotherapy weight loss interventions to prevent obesity-related morbidity and mortality in adults: updated evidence report and systematic review for the us preventive services task force. JAMA 2018 Sep 18;320(11):1172-1191. [doi: 10.1001/jama.2018.7777] [Medline: 30326501]
3. Hales CM, Carroll MD, Fryar CD, Ogden CL. Prevalence of obesity and severe obesity among adults: United States, 2017-2018. NCHS Data Brief 2020 Feb(360):1-8 [FREE Full text] [Medline: 32487284]
4. Ward ZJ, Bleich SN, Cradock AL, Barrett JL, Giles CM, Flax C, et al. Projected U.S. state-level prevalence of adult obesity and severe obesity. N Engl J Med 2019 Dec 19;381(25):2440-2450. [doi: 10.1056/NEJMsa1909301] [Medline: 31851800]
5. 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]
6. Franz MJ, VanWormer JJ, Crain AL, Boucher JL, Histon T, Caplan W, et al. Weight-loss outcomes: a systematic review and meta-analysis of weight-loss clinical trials with a minimum 1-year follow-up. J Am Diet Assoc 2007 Oct;107(10):1755-1767. [doi: 10.1016/j.jada.2007.07.017] [Medline: 17904936]
7. Loveman E, Frampton GK, Shepherd J, Picot J, Cooper K, Bryant J, et al. The clinical effectiveness and cost-effectiveness of long-term weight management schemes for adults: a systematic review. Health Technol Assess 2011 Jan;15(2):1-182 [FREE Full text] [doi: 10.3310/hta15020] [Medline: 21247515]
8. Foley P, Steinberg D, Levine E, Askew S, Batch BC, Puleo EM, et al. Track: A randomized controlled trial of a digital health obesity treatment intervention for medically vulnerable primary care patients. Contemp Clin Trials 2016 Dec;48:12-20 [FREE Full text] [doi: 10.1016/j.cct.2016.03.006] [Medline: 26995281]

9. Weed LL. Medical records that guide and teach. N Engl J Med 1968 Mar 21;278(12):652-7 concl. [doi: 10.1056/NEJM196803212781204] [Medline: 5637250]
10. 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]
11. Burke LE, Conroy MB, Sereika SM, Elci OU, Styn MA, Acharya SD, et al. The effect of electronic self-monitoring on weight loss and dietary intake: a randomized behavioral weight loss trial. Obesity (Silver Spring) 2011 Feb;19(2):338-344 [FREE Full text] [doi: 10.1038/oby.2010.208] [Medline: 20847736]
12. Patel ML, Wakayama LN, Bennett GG. Self-Monitoring via Digital Health in Weight Loss Interventions: A Systematic Review Among Adults with Overweight or Obesity. Obesity (Silver Spring) 2021 Mar;29(3):478-499. [doi: 10.1002/oby.23088] [Medline: 33624440]
13. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
14. Carver CS, Scheier MF. Control theory: a useful conceptual framework for personality-social, clinical, and health psychology. Psychol Bull 1982 Jul;92(1):111-135. [Medline: 7134324]
15. Burke LE, Sereika SM, Music E, Warziski M, Styn MA, Stone A. Using instrumented paper diaries to document self-monitoring patterns in weight loss. Contemp Clin Trials 2008 Mar;29(2):182-193 [FREE Full text] [doi: 10.1016/j.cct.2007.07.004] [Medline: 17702667]
16. Burke LE, Styn MA, Glanz K, Ewing LJ, Elci OU, Conroy MB, et al. SMART trial: A randomized clinical trial of self-monitoring in behavioral weight management-design and baseline findings. Contemp Clin Trials 2009 Nov;30(6):540-551 [FREE Full text] [doi: 10.1016/j.cct.2009.07.003] [Medline: 19665588]
17. Burke LE, Styn MA, Sereika SM, Conroy MB, Ye L, Glanz K, et al. Using mHealth technology to enhance self-monitoring for weight loss: a randomized trial. Am J Prev Med 2012 Jul;43(1):20-26 [FREE Full text] [doi: 10.1016/j.amepre.2012.03.016] [Medline: 22704741]
18. Conroy MB, Yang K, Elci OU, 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]
19. Zheng Y, Terry MA, Danford CA, Ewing LJ, Sereika SM, Goode RW, et al. Experiences of Daily Weighing Among Successful Weight Loss Individuals During a 12-Month Weight Loss Study. West J Nurs Res 2018 Apr;40(4):462-480. [doi: 10.1177/0193945916683399] [Medline: 28322640]
20. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res 2008;10(3):e22 [FREE Full text] [doi: 10.2196/jmir.1030] [Medline: 18725354]
21. Burke LE, Swigart V, Warziski TM, Derro N, Ewing LJ. Experiences of self-monitoring: successes and struggles during treatment for weight loss. Qual Health Res 2009 Jun;19(6):815-828 [FREE Full text] [doi: 10.1177/1049732309335395] [Medline: 19365099]
22. Burke LE, Sereika S, Choo J, Warziski M, Music E, Styn M, et al. Ancillary study to the PREFER trial: a descriptive study of participants' patterns of self-monitoring--rationale, design and preliminary experiences. Contemp Clin Trials 2006 Feb;27(1):23-33. [doi: 10.1016/j.cct.2005.10.006] [Medline: 16311076]
23. Burke LE, Sereika SM, Parmanto B, Beatrice B, Cajita M, Loar I, et al. The SMARTER Trial: Design of a trial testing tailored mHealth feedback to impact self-monitoring of diet, physical activity, and weight. Contemp Clin Trials 2020 Apr;91:105958 [FREE Full text] [doi: 10.1016/j.cct.2020.105958] [Medline: 32087342]
24. Burke LE, Zheng Y, Ma Q, Mancino J, Loar I, Music E, et al. The SMARTER pilot study: Testing feasibility of real-time feedback for dietary self-monitoring. Prev Med Rep 2017 Jun;6:278-285 [FREE Full text] [doi: 10.1016/j.pmedr.2017.03.017] [Medline: 28409090]
25. Turk MW, Elci OU, Wang J, Sereika SM, Ewing LJ, Acharya SD, et al. Self-monitoring as a mediator of weight loss in the SMART randomized clinical trial. Int J Behav Med 2013 Dec;20(4):556-561 [FREE Full text] [doi: 10.1007/s12529-012-9259-9] [Medline: 22936524]
26. 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]
27. Burke LE, Warziski M, Starrett T, Choo J, Music E, Sereika S, et al. Self-monitoring dietary intake: current and future practices. J Ren Nutr 2005 Jul;15(3):281-290. [Medline: 16007557]
28. Burnier M, Schneider MP, Chioléro A, Stubi CL, Brunner HR. Electronic compliance monitoring in resistant hypertension: the basis for rational therapeutic decisions. J Hypertens 2001 Feb;19(2):335-341. [doi: 10.1097/00004872-200102000-00022] [Medline: 11212978]
29. Ambeba EJ, Ye L, Sereika SM, Styn MA, Acharya SD, Sevick MA, et al. The use of mHealth to deliver tailored messages reduces reported energy and fat intake. J Cardiovasc Nurs 2015;30(1):35-43. [doi: 10.1097/JCN.0000000000000120] [Medline: 24434827]
30. Rao G, Kirley K. The future of obesity treatment. JAMA Intern Med 2013 Jan 28;173(2):111. [doi: 10.1001/jamainternmed.2013.1232]

31. Rao G, Kirley K. The future of obesity treatment: comment on "Integrating technology into standard weight loss treatment: a randomized controlled trial". JAMA Intern Med 2013 Jan 28;173(2):111-112. [doi: 10.1001/jamainternmed.2013.1232] [Medline: 23229914]
32. Balbim GM, Marques IG, Marquez DX, Patel D, Sharp LK, Kitsiou S, et al. Using Fitbit as an mHealth Intervention Tool to Promote Physical Activity: Potential Challenges and Solutions. JMIR Mhealth Uhealth 2021 Mar 01;9(3):e25289 [FREE Full text] [doi: 10.2196/25289] [Medline: 33646135]
33. Zheng Y, Klem ML, Sereika SM, Danford CA, Ewing LJ, Burke LE. Self-weighing in weight management: a systematic literature review. Obesity (Silver Spring) 2015 Feb;23(2):256-265. [doi: 10.1002/oby.20946] [Medline: 25521523]
34. Ross KM, Qiu P, You L, Wing RR. Week-to-week predictors of weight loss and regain. Health Psychol 2019 Dec;38(12):1150-1158 [FREE Full text] [doi: 10.1037/hea0000798] [Medline: 31566400]
35. Zheng Y, Burke LE, Danford CA, Ewing LJ, Terry MA, Sereika SM. Patterns of self-weighing behavior and weight change in a weight loss trial. Int J Obes (Lond) 2016 Sep;40(9):1392-1396. [doi: 10.1038/ijo.2016.68] [Medline: 27113642]
36. Burke LE, Styn MA, Sereika SM, Conroy MB, Ye L, Glanz K, et al. Using mHealth technology to enhance self-monitoring for weight loss: a randomized trial. Am J Prev Med 2012 Jul;43(1):20-26 [FREE Full text] [doi: 10.1016/j.amepre.2012.03.016] [Medline: 22704741]
37. Burke LE, Sereika SM, Parmanto B, Bizhanova Z, Kariuki JK, Cheng J, et al. Effect of tailored, daily feedback with lifestyle self-monitoring on weight loss: The SMARTER randomized clinical trial. Obesity (Silver Spring) 2022 Jan;30(1):75-84. [doi: 10.1002/oby.23321] [Medline: 34898011]
38. Jakicic JM, Clark K, Coleman E, Donnelly JE, Foreyt J, Melanson E, American College of Sports Medicine. American College of Sports Medicine position stand. Appropriate intervention strategies for weight loss and prevention of weight regain for adults. Med Sci Sports Exerc 2001 Dec;33(12):2145-2156. [doi: 10.1097/00005768-200112000-00026] [Medline: 11740312]
39. Wadden TA, Berkowitz RI, Sarwer DB, Prus-Wisniewski R, Steinberg C. Benefits of lifestyle modification in the pharmacologic treatment of obesity: a randomized trial. Arch Intern Med 2001 Jan 22;161(2):218-227. [doi: 10.1001/archinte.161.2.218] [Medline: 11176735]
40. Thomas JG, Bond DS, Raynor HA, Papandonatos GD, Wing RR. Comparison of Smartphone-Based Behavioral Obesity Treatment With Gold Standard Group Treatment and Control: A Randomized Trial. Obesity (Silver Spring) 2019 Apr;27(4):572-580 [FREE Full text] [doi: 10.1002/oby.22410] [Medline: 30779333]
41. Franz MJ, VanWormer JJ, Crain AL, Boucher JL, Histon T, Caplan W, et al. Weight-loss outcomes: a systematic review and meta-analysis of weight-loss clinical trials with a minimum 1-year follow-up. J Am Diet Assoc 2007 Oct;107(10):1755-1767. [doi: 10.1016/j.jada.2007.07.017] [Medline: 17904936]
42. Spring B, Pellegrini CA, Pfammatter A, Duncan JM, Pictor A, McFadden HG, et al. Effects of an abbreviated obesity intervention supported by mobile technology: The ENGAGED randomized clinical trial. Obesity (Silver Spring) 2017 Jul;25(7):1191-1198 [FREE Full text] [doi: 10.1002/oby.21842] [Medline: 28494136]
43. Patel ML, Hopkins CM, Brooks TL, Bennett GG. Comparing Self-Monitoring Strategies for Weight Loss in a Smartphone App: Randomized Controlled Trial. JMIR Mhealth Uhealth 2019 Feb 28;7(2):e12209 [FREE Full text] [doi: 10.2196/12209] [Medline: 30816851]
44. Turner-McGrievy G, Tate D. Tweets, Apps, and Pods: Results of the 6-month Mobile Pounds Off Digitally (Mobile POD) randomized weight-loss intervention among adults. J Med Internet Res 2011 Dec 20;13(4):e120 [FREE Full text] [doi: 10.2196/jmir.1841] [Medline: 22186428]
45. Block G, Azar KM, Romanelli RJ, Block TJ, Hopkins D, Carpenter HA, et al. Diabetes Prevention and Weight Loss with a Fully Automated Behavioral Intervention by Email, Web, and Mobile Phone: A Randomized Controlled Trial Among Persons with Prediabetes. J Med Internet Res 2015 Oct 23;17(10):e240 [FREE Full text] [doi: 10.2196/jmir.4897] [Medline: 26499966]
46. Littman AJ, Boyko EJ, McDonell MB, Fihn SD. Evaluation of a weight management program for veterans. Prev Chronic Dis 2012;9:E99 [FREE Full text] [doi: 10.5888/pcd9.110267] [Medline: 22595323]
47. Glasgow RE, Nelson CC, Kearney KA, Reid R, Ritzwoller DP, Strecher VJ, et al. Reach, engagement, and retention in an Internet-based weight loss program in a multi-site randomized controlled trial. J Med Internet Res 2007 May 09;9(2):e11 [FREE Full text] [doi: 10.2196/jmir.9.2.e11] [Medline: 17513282]
48. 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]
49. Amagai S, Pila S, Kaat AJ, Nowinski CJ, Gershon RC. Challenges in Participant Engagement and Retention Using Mobile Health Apps: Literature Review. J Med Internet Res 2022 Apr 26;24(4):e35120 [FREE Full text] [doi: 10.2196/35120] [Medline: 35471414]
50. Tate DF, Wing RR, Winett RA. Using Internet technology to deliver a behavioral weight loss program. JAMA 2001 Mar 07;285(9):1172-1177. [doi: 10.1001/jama.285.9.1172] [Medline: 11231746]
51. West D, Krukowski R, Stansbury M, Ogden D, Borden J, Harvey J. Examining weekly facilitated group sessions and counselor - crafted self - monitoring feedback on treatment outcome in digital weight control: A pilot factorial study. Obesity Science & Practice 2022 Jan 05:A [FREE Full text] [doi: 10.1002/osp4.585]

52. DiClemente CC, Marinilli AS, Singh M, Bellino LE. The role of feedback in the process of health behavior change. Am J Health Behav 2001;25(3):217-227. [doi: 10.5993/ajhb.25.3.8] [Medline: 11322620]
53. Bidargaddi N, Almirall D, Murphy S, Nahum-Shani I, Kovalcik M, Pituch T, et al. To Prompt or Not to Prompt? A Microrandomized Trial of Time-Varying Push Notifications to Increase Proximal Engagement With a Mobile Health App. JMIR Mhealth Uhealth 2018 Nov 29;6(11):e10123 [FREE Full text] [doi: 10.2196/10123] [Medline: 30497999]
54. Hutchesson MJ, Tan CY, Morgan P, Callister R, Collins C. Enhancement of Self-Monitoring in a Web-Based Weight Loss Program by Extra Individualized Feedback and Reminders: Randomized Trial. J Med Internet Res 2016 Apr 12;18(4):e82 [FREE Full text] [doi: 10.2196/jmir.4100] [Medline: 27072817]
55. Bell K, Onyeukwu C, McClincy M, Allen M, Bechard L, Mukherjee A, et al. Verification of a Portable Motion Tracking System for Remote Management of Physical Rehabilitation of the Knee. Sensors (Basel) 2019 Feb 28;19(5):1021 [FREE Full text] [doi: 10.3390/s19051021] [Medline: 30823373]
56. Lyzwinski LN. A systematic review and meta-analysis of mobile devices and weight loss with an intervention content analysis. J Pers Med 2014 Jun 30;4(3):311-385 [FREE Full text] [doi: 10.3390/jpm4030311] [Medline: 25563356]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
FB: feedback
NIH: National Institutes of Health
PA: physical activity
SBT: standard behavioral treatment
SM: self-monitoring

Edited by A Mavragani; submitted 01.04.22; peer-reviewed by M Patel, G Papandonatos; comments to author 28.04.22; revised version received 19.05.22; accepted 20.05.22; published 05.07.22.

Please cite as:
Burke LE, Sereika SM, Bizhanova Z, Parmanto B, Kariuki J, Cheng J, Beatrice B, Cedillo M, Pulantara IW, Wang Y, Loar I, Conroy MB
The Effect of Tailored, Daily, Smartphone Feedback to Lifestyle Self-Monitoring on Weight Loss at 12 Months: the SMARTER Randomized Clinical Trial
J Med Internet Res 2022;24(7):e38243
URL: https://www.jmir.org/2022/7/e38243
doi:10.2196/38243
PMID:35787516

Original Paper

The Effect of a Future-Self Avatar Mobile Health Intervention (FutureMe) on Physical Activity and Food Purchases: Randomized Controlled Trial

Annette Mönninghoff[1,2], PhD; Klaus Fuchs[3], PhD; Jing Wu[4], MSc; Jan Albert[4], BASc; Simon Mayer[4], Prof Dr

[1]Institute for Mobility, University of St. Gallen, St Gallen, Switzerland

[2]Institute for Customer Insight, University of St. Gallen, St Gallen, Switzerland

[3]ETH AI Center, ETH Zurich, Zurich, Switzerland

[4]Institute for Computer Science, University of St. Gallen, St Gallen, Switzerland

Corresponding Author:
Annette Mönninghoff, PhD
Institute for Mobility
University of St. Gallen
Bahnhofstrasse 8
St Gallen, 9000
Switzerland
Phone: 41 712242132
Email: Annette.Moenninghoff@unisg.ch

Abstract

Background: Insufficient physical activity and unhealthy diets are contributing to the rise in noncommunicable diseases. Preventative mobile health (mHealth) interventions may help reverse this trend, but present bias might reduce their effectiveness. Future-self avatar interventions have resulted in behavior change in related fields, yet evidence of whether such interventions can change health behavior is lacking.

Objective: We aimed to investigate the impact of a future-self avatar mHealth intervention on physical activity and food purchasing behavior and examine the feasibility of a novel automated nutrition tracking system. We also aimed to understand how this intervention impacts related attitudinal and motivational constructs.

Methods: We conducted a 12-week parallel randomized controlled trial (RCT), followed by semistructured interviews. German-speaking smartphone users aged ≥18 years living in Switzerland and using at least one of the two leading Swiss grocery loyalty cards, were recruited for the trial. Data were collected from November 2020 to April 2021. The intervention group received the FutureMe intervention, a physical activity and food purchase tracking mobile phone app that uses a future-self avatar as the primary interface and provides participants with personalized food basket analysis and shopping tips. The control group received a conventional text- and graphic-based primary interface intervention. We pioneered a novel system to track nutrition by leveraging digital receipts from loyalty card data and analyzing food purchases in a fully automated way. Data were consolidated in 4-week intervals, and nonparametric tests were conducted to test for within- and between-group differences.

Results: We recruited 167 participants, and 95 eligible participants were randomized into either the intervention (n=42) or control group (n=53). The median age was 44 years (IQR 19), and the gender ratio was balanced (female 52/95, 55%). Attrition was unexpectedly high with only 30 participants completing the intervention, negatively impacting the statistical power. The FutureMe intervention led to small statistically insignificant increases in physical activity (median +242 steps/day) and small insignificant improvements in the nutritional quality of food purchases (median −1.28 British Food Standards Agency Nutrient Profiling System Dietary Index points) at the end of the intervention. Intrinsic motivation significantly increased (P=.03) in the FutureMe group, but decreased in the control group. Outcome expectancy directionally increased in the FutureMe group, but decreased in the control group. Leveraging loyalty card data to track the nutritional quality of food purchases was found to be a feasible and accepted fully automated nutrition tracking system.

Conclusions: Preventative future-self avatar mHealth interventions promise to encourage improvements in physical activity and food purchasing behavior in healthy population groups. A full-powered RCT is needed to confirm this preliminary evidence and to investigate how future-self avatars might be modified to reduce attrition, overcome present bias, and promote sustainable behavior change.

Trial Registration: ClinicalTrials.gov NCT04505124; https://clinicaltrials.gov/ct2/show/NCT04505124

(J Med Internet Res 2022;24(7):e32487) doi:10.2196/32487

KEYWORDS

mHealth; mobile health; preventative medicine; avatar; present bias; nutrition tracking; physical activity; randomized controlled trial

Introduction

Noncommunicable Diseases and Risk Factors

Noncommunicable diseases (NCDs) are today's leading causes of death. They are responsible for over half of the years lived in disability and drive 59%-80% of total health care costs in affluent developed countries like New Zealand and Switzerland [1-3]. Among others, insufficient physical activity and an unhealthy sodium-rich diet are recognized as key risk factors for NCDs by the World Health Organization (WHO) [4]. Increased salt consumption has been shown to increase blood pressure, while sodium-rich, low-fruit, and low–whole-grain diets may increase the risk of death from cardiovascular diseases, cancer, and diabetes [5-7]. Insufficient physical activity has been linked to increased risks of cancer, cardiovascular diseases, diabetes, dementia, and depression [8-11]. To combat the rise in NCDs, the WHO has released a global action plan providing clear recommendations on the required physical activity levels and maximum salt consumption [4]. However, the latest studies showed that globally, 28% of adults do not meet the WHO's physical activity guidelines [12]. Moreover, salt intake was approximately twice the recommended amount in 2010 [13], and has not improved significantly since [14].

Physical Activity and Salt Reduction Mobile Health Interventions

Wide-scale prevention programs are needed to make significant improvements toward realizing the WHO's NCD action plan. Mobile health (mHealth) interventions can make an important contribution toward efficiently reaching the wider population. While numerous physical activity mHealth interventions have proven effective in a wide range of populations [15-18], their effectiveness in decreasing salt consumption is less clear [19]. Preliminary results suggest that while mHealth interventions might lower participants' salt consumption by raising awareness, the quality of the evidence gathered in such studies is very low. Scholars have also criticized the lack of rigorous research designs [19]. In addition, mHealth nutrition interventions, such as those aiming to reduce sodium intake, regularly experience low user adoption and high attrition [20-22]. A review of commercially available nutrition-tracking mHealth apps showed that most such interventions relied on manual or semiautomated tracking technologies [23], and required significant time investment from users, favoring underreporting [24] and high attrition rates as a result. There is a lack of studies employing improved technological means, such as automated digital receipt processing [25,26], image analysis, and sensor-based tracking. Thus, there is a need for high-quality research that evaluates automated tracking systems in rigorous research settings [23,27].

Challenges of Preventative mHealth Programs

Preventative mHealth interventions (ie, mHealth interventions targeting healthy population groups) have been shown to be less effective than curative mHealth interventions (ie, mHealth interventions targeting sick or at-risk population groups) [15,18]. The reasons behind this moderating effect are manifold and remain little explored. The Health Action Process Approach (HAPA) is a widely used and validated theoretical framework that can serve as a starting point for explaining differences in intervention effectiveness [28]. According to the associated model, health behavior change is best described as a 2-phase process consisting of a preintentional motivational phase and a postintentional volitional phase [28]. Within the first phase, risk awareness (ie, the awareness that one's behavior may cause a health risk), positive outcome expectancies (ie, the belief that a change in behavior will lead to reduced health risk), and action self-efficacy (ie, the belief in one's ability to perform preventative health behavior) are fundamental for intention formation [28]. However, present bias may cause low risk awareness and decrease positive outcome expectancies in healthy population groups and thus limit the effectiveness of preventative mHealth interventions [29]. Present bias describes a situation where people take suboptimal decisions as they over-discount future gains [29-32]. When facing preventative health decisions, consumers are asked to make unpleasant behavior changes today (ie, eating more vegetables instead of meat and exercising instead of watching movies), while the benefits of their behaviors occur in the future (ie, better future health and longer life expectancy). In reverse, the negative consequences of a risky lifestyle are not immediately evident but only occur with a time lag, hindering the formation of positive outcome expectancies (ie, low symptom salience) [33].

Future-Self Avatars to Overcome Present Bias and Increase Risk Awareness

Existing research argues that immediate monetary incentives can help overcome present bias in health decision-making and improve adaptation of preventative mHealth interventions [34-36]. However, extrinsic monetary incentives might lead to motivational crowding-out and diminish intrinsic motivation for behavior change over time [37,38]. Alternative strategies to overcome present bias and to increase risk awareness have been developed in related fields, such as pension planning [31] and taxation [39]. A little explored possible solution is the use of future-self avatars (ie, avatars that confront users with an aged version of themselves). Such avatars have been shown to increase users' tendency to accept later gains over immediate ones [31]. Evidence on the impact of avatars on health decision-making is still scarce, but preliminary results indicate that avatars can be leveraged to increase physical activity in obese populations [40], to promote smoking cessation in young

adults [41], to reduce sodium intake [42], and to help manage depression [43]. To date, however, no study has investigated the impact of a preventative future-self avatar mHealth intervention in a randomized controlled trial (RCT) setting.

Objectives

Accordingly, the primary aim of this study was 2 fold. First, the study aimed to understand the impact of a preventative future-self avatar mHealth intervention on physical activity and on the nutritional quality of food purchases. Second, it aimed to test the acceptability of a fully automated nutrition tracking system leveraging grocery receipts in an RCT. Our secondary aim was to understand the impact of a future-self avatar preventative mHealth intervention on motivational and recovery self-efficacy, outcome expectancy, and types of motivation.

Methods

Study Design

We conducted our study in cooperation with a large Swiss health insurance company that facilitated participant recruitment. A 12-week parallel RCT design was employed, followed by semistructured interviews with a selection of 15 participants who completed the trial. Recruitment began in November 2020 and ended in December 2020. Data were collected from November 2020 to April 2021. Potential participants were contacted via the insurance company's email newsletter and informed about the study, eligibility criteria, and data privacy policies. Additionally, social media networks were leveraged to recruit trial participants. Potential participants received a direct link to the iOS or Android app store where they could download the app-based intervention. After downloading the mobile app, answering a survey on eligibility criteria, and consenting to the data privacy statement, participants were randomized into an intervention or a control group (1:1 ratio) via a random allocation algorithm programmed into the app. Participants did not receive any information about the existence of 2 different app versions (intervention and control). Apart from the in-depth interviews, data collection was fully automated, thus preventing reporting, detection, and performance bias. The semistructured interviews were conducted by trained researchers who were neither the primary researchers nor the authors of this study. Selection of interview participants was blinded, meaning that interviewers had no knowledge of the interviewees' group allocation or the outcome data collected in the RCT. The trial has been registered on ClinicalTrials.gov (NCT04505124), and the study has been reported in accordance with the CONSORT guidelines [44].

Study Population

Eligible participants were smartphone users aged ≥18 years, living in Switzerland, and German-speaking. Participants had to be participating in at least one of the two leading grocery loyalty card programs in Switzerland (Multimedia Appendix 1). They were asked to confirm that they had no medical condition that prevented increased levels of physical activity or changes in daily nutrition.

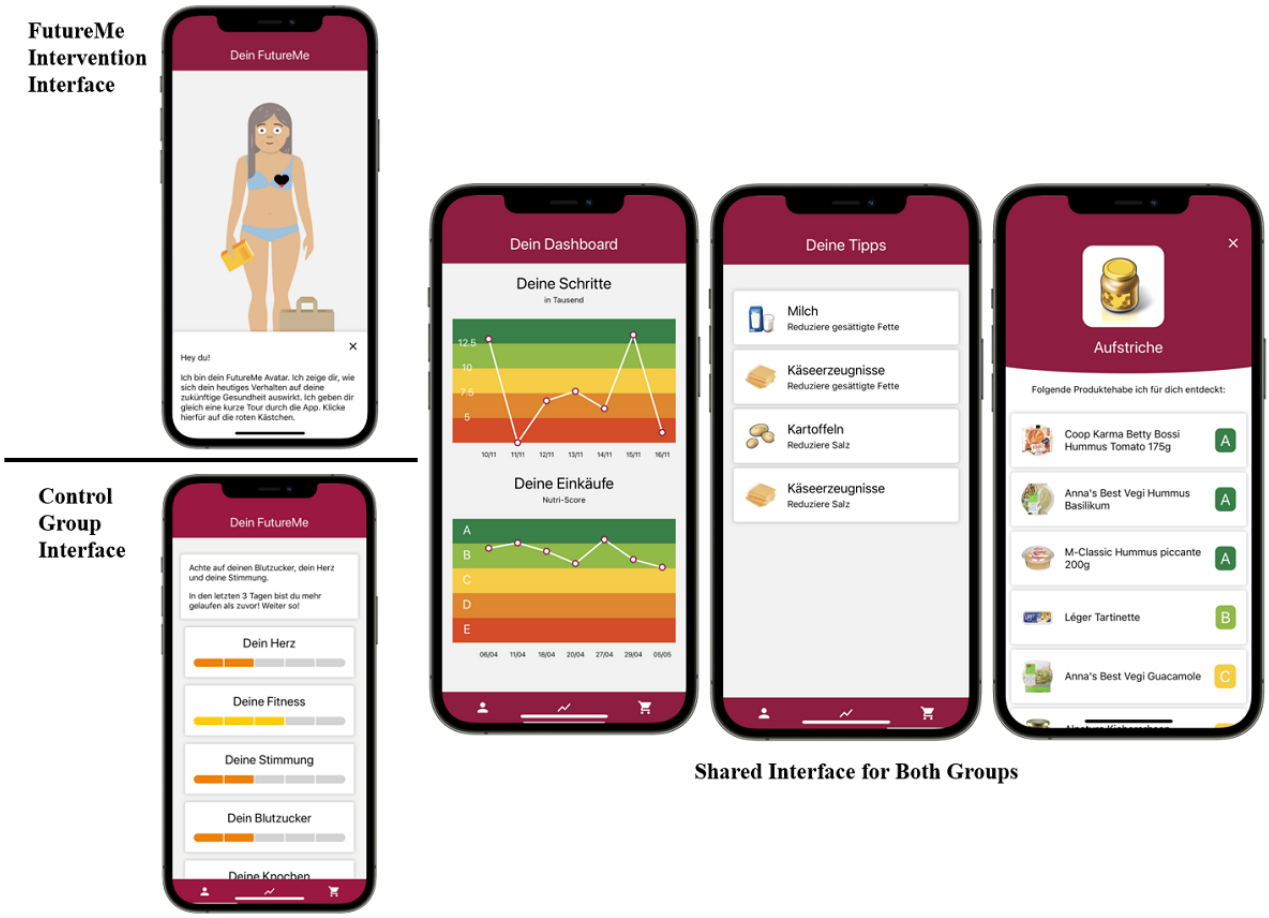
FutureMe Intervention

The FutureMe app aims to promote the overall nutritional quality of participants' food purchases and to encourage increases in physical activity to ≥7500 steps per day, based on prior findings that health benefits can be achieved at this level [45-47]. Participants allocated to the FutureMe intervention received the FutureMe app. This comprised a future-self avatar, a dashboard displaying the number of daily steps and the nutritional quality of food purchases, a personalized basket analysis showing food categories with the highest improvement potentials, and personalized product recommendations of healthier product alternatives within the food categories purchased by the participant (Figure 1). All app components were updated daily, ensuring that participants received the most recent data. The intervention was fully automated and required no human-to-human interaction, making it highly scalable, tailored, and cost-effective [15].

The personalized basket analysis and product recommendations were based on the British Food Standards Agency Nutrient Profiling System Dietary Index (FSA-NPS DI), which is also referred to as Nutri-Score [48]. The intervention mechanism was co-designed with the Swiss Society for Nutrition and aimed to induce a behavior shift toward healthier choices within the frequently purchased categories that contributed most negatively to a user's recent FSA-NPS DI assessment. Concretely, in the FutureMe intervention, we evaluated the 7 most recently purchased baskets using FSA-NPS DI points and assessed the contribution of negative points (ie, sugar, sodium, saturated fats, and energy density) by each food category. Specifically, we calculated the weight-averaged contribution of each food category across the 4 negative FSA-NPS DI dimensions. We identified up to four food categories (out of the 125 food categories used in the food composition database of the study) that contributed most negatively to a user's FSA-NPS DI score. The FutureMe app then recommended healthier product alternatives within these problematic categories. The presented product recommendations were meant to help users make healthier choices by substituting frequently consumed and rather unhealthy food items with alternatives of higher nutritional quality, without requiring users to modify the structure of their diets or key dietary habits (eg, recommending lower-salt cheese alternatives). In addition to showing healthier product alternatives, users received relevant nutritional tips that were adapted based on the identified problematic food categories.

Although no scholarly consensus toward a single metric to capture the nutritional quality of food purchases exists [48], the Nutri-Score framework and the underlying established FSA-NPS DI system allow for a reliable heuristic to assess the improvement potentials of purchased grocery baskets, thus enabling automated personalized purchase recommendations for a specific user.

Figure 1. Overview of the FutureMe and control intervention mobile apps.



We selected the Nutri-Score framework as a monitoring proxy for the nutritional quality of recently purchased baskets and as an intervention mechanism to support healthier food choices within the FutureMe app for multiple reasons. First, the FSA-NPS DI has been identified as a useful and validated tool to discriminate individuals according to the quality of the diet, allowing the monitoring of dietary change [49]. Second, within a comparative study among purchase quality indicators in Switzerland, the Nutri-Score framework has been validated to estimate diet activities from shopping data better than other purchase indicators [50]. Third, when compared to other front-of-package labels, the Nutri-Score label has been validated to show superior effectiveness among consumers, especially among the at-risk population [51]. Fourth, although the Nutri-Score framework was developed for rating individual products and not aggregated shopping baskets, its scoring mechanism has been proven to correlate with overall dietary behavior when applied on grocery purchase data sets [50].

To the best of our knowledge, although the FSA-NPS DI and Nutri-Score frameworks are established concepts, applying them within a smartphone-mediated intervention design using digital receipts from loyalty card data has not been done before.

Future-Self Avatar

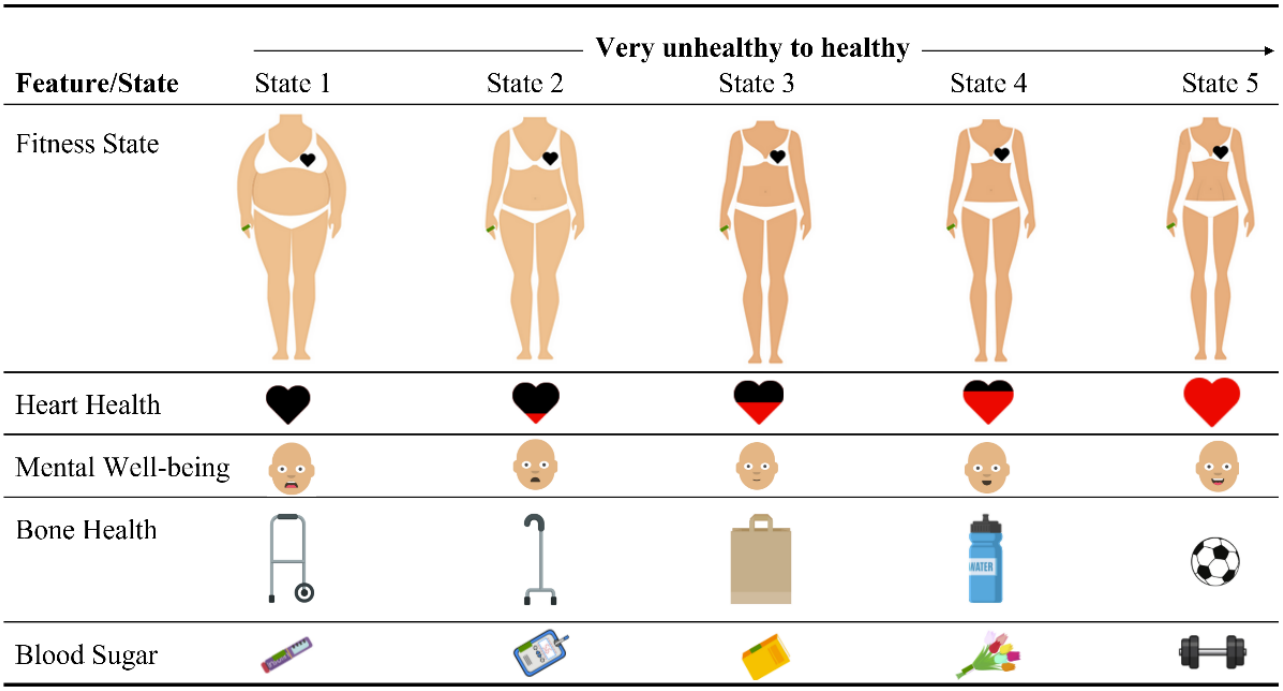
The future-self avatar was designed based on the HAPA behavior change model and encouraging findings of a prior pilot study by Fuchs et al [42]. The future-self avatar aims to increase participants' awareness of the future health consequences of their current nutritional and activity behaviors in a fun and engaging way (cf, risk perception, outcome expectancy, and self-efficacy [28]). It thus seeks to minimize the negative impact of present bias [31]. Participants could personalize their avatar during the sign-up process to ensure that the avatar best resembled themselves, in order to increase personal relevance and to positively impact health behavior change [52]. The avatar was depicted using a cartoon-style unanimated 2D design to avoid feelings of eeriness and repulsion as described by the uncanny valley effect [53]. When first opening the app, participants were exposed to an aging simulation in which their avatar aged +20 years compared to their current age. The future-self avatar had 5 features that consumers could affect with their food purchases and physical activity behavior, with each feature having 5 different states (see Table 1 and Figure 2). As the nutritional recommendations of our research were based on the FSA score framework [53], the avatar health features captured 6 nutritional subcategories of this framework. The health features were based on scientific literature but were also designed to be easily understood by the broad group of trial participants. The calculation scheme for each state is provided in Multimedia Appendix 2.

Table 1. System rules defining the feature states of the future-self avatar.

Feature and behavioral influencer	Weighting	Calculation scheme	Unit
Fitness state			
Physical activity	100%	Average past 7 days	Steps/day
Heart health			
Sodium	50%	Average past 12 baskets	FSA-NPS DI[a] points
Saturated fatty acids	50%	Average past 12 baskets	FSA-NPS DI points
Mental well-being			
Fruits and vegetables	50%	Average past 12 baskets	FSA-NPS DI points
Fiber	50%	Average past 12 baskets	FSA-NPS DI points
Bone health			
Protein	100%	Average past 12 baskets	FSA-NPS DI points
Blood sugar			
Sugar	100%	Average past 12 baskets	FSA-NPS DI points

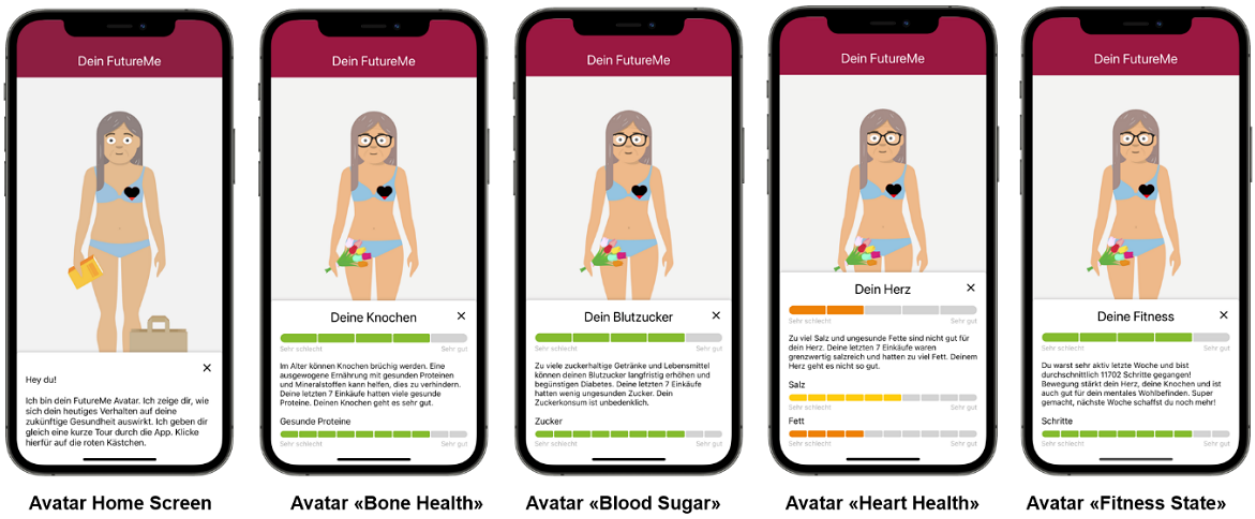
[a]FSA-NPS DI: British Food Standards Agency Nutrient Profiling System Dietary Index [53].

Figure 2. Avatar feature states.



The 5 features were clickable. If clicked, a pop-up screen opened, providing users with consequential health feedback and how their individual past behavior had impacted the feature. Participants were provided with an encouraging message to improve the respective health behavior (eg, “You were very active last week and took an average 11,702 steps/day! Physical activity improves your heart health, strengthens your bones, and also benefits your mental well-being! Well done, next week you can do even more” for *Fitness State*; Figure 3).

Figure 3. Overview of avatar features.



Fully Automated Nutrition and Physical Activity Tracking

Nutrition and physical activity tracking within the FutureMe intervention app pioneered a novel and fully automated tracking system. Food purchases recorded via loyalty cards were used as a proxy for nutrition, as South African studies leveraging credit card data have shown this to be a promising method for changing everyday health behavior [25,26]. We gathered food purchase data using the loyalty cards from Switzerland's 2 leading grocery retailers. Data were analyzed using a comprehensive product database comprising over 55,000 products and their nutritional information. Participants were requested to connect their loyalty cards once while signing up. Step data were collected by the accelerometer on participants' mobile phones.

Control Intervention

Apart from the future-self avatar, the control intervention was identical to the FutureMe intervention (see Figure 1). Instead of the future-self avatar, the control app used a more traditional text- and graphics-based interface to provide consequential health feedback to participants (see Figure 1). The control intervention used the same automated data tracking system as the FutureMe intervention.

Outcomes and Measurement

Demographic information as well as attitudinal and motivational constructs were collected at baseline (T0), and after completing the 12-week intervention (T3), using an online survey that was programmed into the mHealth app. Step and food purchase data were collected at baseline (T0) and continuously throughout the 12-week intervention period using the fully automated tracking system previously described. At baseline, we collected steps per day for the past 6 days prior to sign-up and data for all grocery baskets of the past 2 years that were registered on the loyalty cards. User engagement was recorded continuously for 12 weeks after completing the sign-up process. Continuously collected data were aggregated into three 4-week periods (T1, T2, and T3). Qualitative data were collected from a group of randomly selected participants in semistructured video interviews after the 12-week intervention period (T3).

The primary outcomes were as follows:

1. Physical activity (steps/day) was measured objectively to avoid reporting biases [54]. Step information was collected through the built-in accelerometer in iOS or Android mobile phones, as these devices are widely owned and require no additional investment [55], and thus offer a scalable solution to track physical activity. Mobile phones have been shown to reliably measure steps if carried continuously by participants and when steps are taken at a modest to fast pace [55-57].
2. The nutritional quality of food purchases was measured based on the British Food Standards Agency Nutrient Profiling System [49,50] (in FSA-NPS DI points; –15 most healthy to +40 least healthy) [58]. Nutritional quality was calculated by separating solid foods and beverages, in accordance with the calculation method suggested by Julia et al [59]. We only report results for solid foods.

The secondary outcomes were the nutritional subcategories comprising the Nutri-Score framework (secondary outcomes 1-6) [60], user engagement (secondary outcome 7), and attitudinal and motivational constructs (secondary outcomes 8-12) as follows:

1. Sugars, excluding fructose and lactose, in grams (g) per 100 g food purchases (in FSA-NPS DI points; 0 most healthy to +10 least healthy).
2. Saturated fatty acids in g per 100 g food purchases (in FSA-NPS DI points; 0 most healthy to +10 least healthy).
3. Sodium in mg per 100 g food purchases (in FSA-NPS DI points; 0 most healthy to +10 least healthy).
4. Fruits, vegetables, legumes, and nuts in % per 100 g food purchases (in FSA-NPS DI points; 0 least healthy to +5 most healthy).
5. Fiber in g per 100 g food purchases (in FSA-NPS DI points; 0 least healthy to +5 most healthy).

6. Protein in g per 100 g food purchases (in FSA-NPS DI points; 0 least healthy to +5 most healthy).
7. User engagement (app logins/week).
8. Motivational self-efficacy, which was measured with an adjusted 4-item scale adopted from Schwarzer et al [61] on a 7-point Likert scale (1 [completely disagree] to 7 [completely agree]). The scale measured participants' confidence in their capability to be physically active and shop healthily in general (eg, "In my everyday life, I know how to shop healthily") (Multimedia Appendix 3). With regard to internal consistency, Cronbach α was .728.
9. Recovery self-efficacy, which was measured with an adjusted 2-item scale adopted from Schwarzer et al [61] on a 7-point Likert scale (1 [completely disagree] to 7 [completely agree]). The scale measured participants' confidence in their capability to be physically active and shop healthily after a setback (eg, "After my vacation, I'm sure that I'll go back to balanced shopping, even if I have to get used to it again") (Multimedia Appendix 3). With regard to internal consistency, Cronbach α was .652.
10. Outcome expectancy, which was measured with an adjusted 6-item outcome expectancy scale adopted from Renner and Schwarzer [62] on a 7-point Likert scale (1 [completely disagree] to 7 [completely agree]) that measured participants' expectations about how their present behavior will impact their future health (eg, "I believe that I can positively influence my health in old age with my current exercise and shopping behavior") (Multimedia Appendix 3). With regard to internal consistency, Cronbach α was .934.
11. Intrinsic motivation, which was measured with an adjusted 3-item autonomous motivation scale [63] on a 7-point Likert scale (1 [completely disagree] to 7 [completely agree]). The scale measured the degree to which participants were intrinsically motivated to be physically active and shop healthily (eg, "I exercise and shop healthily because I personally believe it's best for my health") (Multimedia Appendix 3). With regard to internal consistency, Cronbach α was .838.
12. Extrinsic motivation, which was measured with an adjusted 3-item controlled motivation scale [63] on a 7-point Likert scale (1 [completely disagree] to 7 [completely agree]). The scale measured the degree to which participants were extrinsically motivated to be physically active and shop healthily (eg, "I exercise and shop healthily because I want to see positive metrics on my activity tracker and health app") (Multimedia Appendix 3). With regard to internal consistency, Cronbach α was .690.

In addition to the quantitative outcomes, we collected qualitative data in 15 semistructured video interviews. Data were gathered on app usefulness and ease of use, as these factors have been shown to impact technology acceptance [64]. Furthermore, data were collected on the relationship that participants formed with the avatar and on areas for improving the FutureMe intervention in order to guide future research.

Sample Size

No other studies have so far compared a future-self avatar interface with a more conventional self-monitoring interface regarding the ability to motivate both physical activity increases and nutrition improvements in a healthy population sample. Given the limited availability of mHealth studies reporting Nutri-Score improvements based on shopping data, we focused on results from comparable mHealth physical activity studies to calculate the sample size. A systematic review of mHealth physical activity interventions found the step increase at the end of the intervention in a mixed population sample at 1566 steps [15] compared with both active and passive control groups. Baseline activity measurements for healthy population samples have been reported to range from 6745 to 6994 (SD 2422 to 2620) steps/day [65,66]. Using an α of .05 with a power of 80%, we estimated the minimum sample size for our trial to be 74-88 participants [67]. Dropout rates in mHealth studies vary significantly, depending on intervention duration and intervention components. A comparable 12-week mHealth physical activity study using scalable intervention components without human-to-human interaction reported a dropout rate of 22.4% [68]. Anticipating a 20% dropout rate, we inflated the sample size to 100.

Statistical Analysis

Nonparametric tests were conducted to examine intervention effects. A Wilcoxon signed-rank test was used to compare the outcome variables before and after the intervention to test for within-group effects. To test for between-group differences between the FutureMe and control groups, we conducted Mann-Whitney *U* tests or Pearson chi-square tests. The level of significance was .05. Analyses were performed with SPSS Statistics 27 (IBM Corp) and Python 3.8.5 (Python Software Foundation).

Continuously measured data (physical activity, nutritional quality of food purchases, nutritional subcategories, and user engagement) were analyzed in 4-week periods, with T1 representing the mean value summarizing weeks 1-4, T2 summarizing weeks 5-8, and T3 summarizing weeks 9-12. Baseline values (T0) were also summarized depending on the outcome variable and data availability. For physical activity, T0 represents the mean steps per day of the 6 days prior to trial commencement. For the nutritional quality of food purchases and the nutritional subcategories, T0 represents all foods purchased within 4 weeks prior to trial commencement.

Sensitivity analysis was conducted for demographic variables and primary outcomes to test for randomness of missing data and attrition bias [69,70]. Missing data were not replaced given the limitations of imputation methods in samples with high attrition where data are missing at random [69]. To test for intervention effects and statistical differences, we used complete case analysis.

Qualitative Data Analysis

All 15 semistructured video interviews were audio recorded, fully transcribed, coded, and analyzed following the principles of thematic content analysis [71]. Additionally, demographic information on all 15 interview candidates was quantitatively summarized.

Ethics Approval

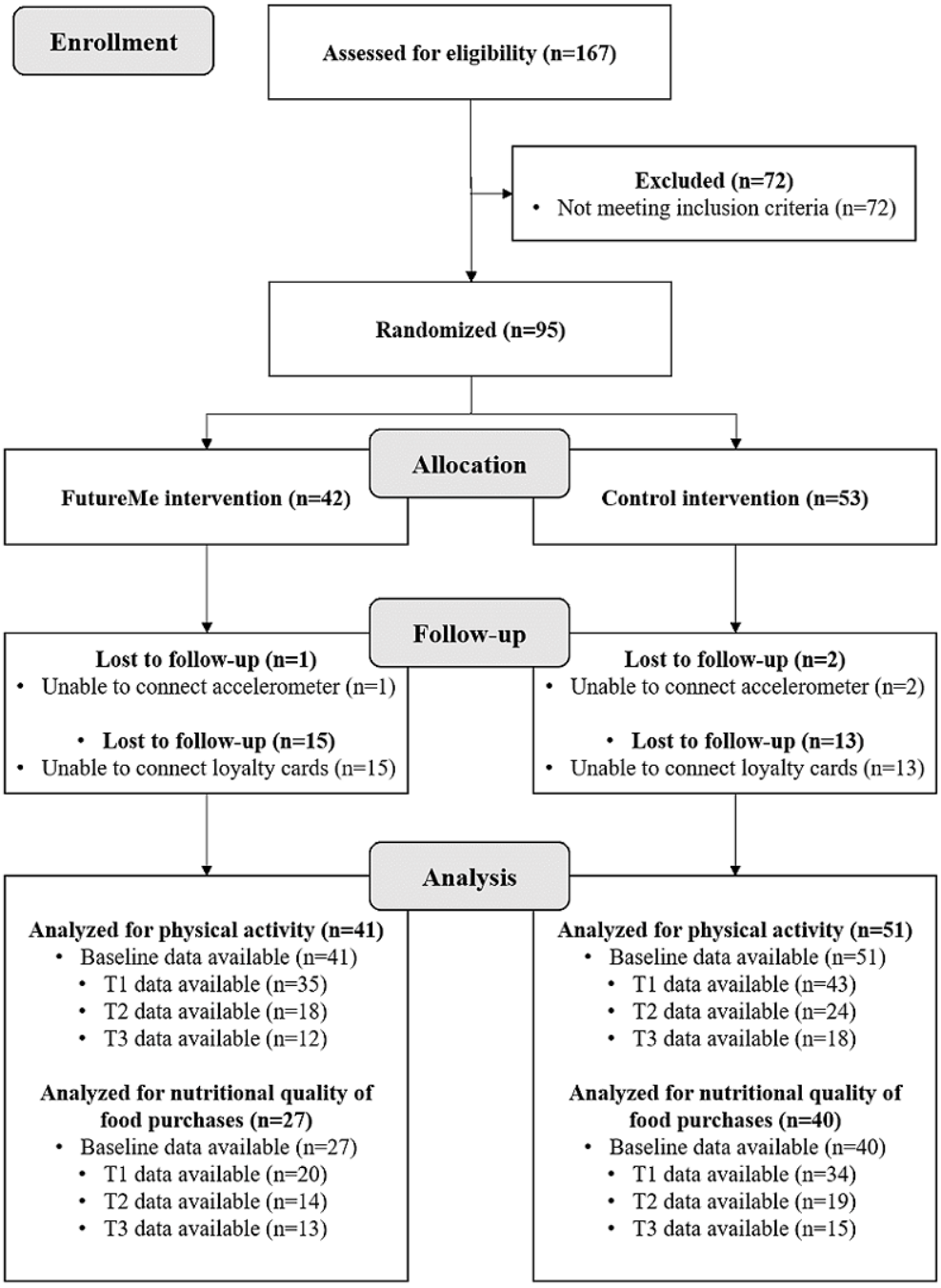
Participants gave their electronic informed consent to participate in the study prior to commencement. The Ethics Committee of the University of St. Gallen approved the protocol, recruitment strategies, and data privacy policies of this study (HSG-EC-2020-06-12-A).

Results

Demographic Data and Baseline Characteristics

As shown in the CONSORT flow chart (Figure 4), 167 participants downloaded the FutureMe/Control app from November 11 to December 15, 2020. Seventy-two participants were excluded as they did not meet the inclusion criteria. Ninety-five participants were randomized to either the FutureMe intervention (n=42) or the control intervention (n=53).

Figure 4. CONSORT flow chart. Limited data were available at T1-3 as some participants from both groups stopped using the FutureMe/control app. Data were only automatically pulled when users logged into the app.



One participant in the intervention group and 2 participants in the control group were lost as they were unable to connect their mobile phone accelerometer to the intervention/control app. Fifteen participants in the intervention group and 13 participants in the control group were lost as they were unable to connect their loyalty cards to the FutureMe app. We saw unexpectedly high attrition during the 12-week intervention period (n=95, 68.4%), with only 13 participants from the intervention group and 18 participants from the control group using the app during the last 4 weeks of the trial.

As shown in Table 2, the median age of the participants was 44 years (IQR 19). The gender ratio was balanced, with 55% (52/95) female participants. The median household size was 2 (IQR 2), and most participants had normal body weight (56/95, 59%). The majority (76/95, 80%) of participants were recruited into the trial through the online panel of a Swiss health insurance company. No baseline differences existed between the intervention and control groups with regard to demographic data.

Table 2. Demographic data at baseline.

Characteristic	Total population (N=95)	Intervention group (n=42)	Control group (n=53)	*P* value
Gender, n (%)				>.99[a]
Male	43 (45)	19 (45)	24 (45)	
Female	52 (55)	23 (55)	29 (55)	
Age (years), median (IQR)	44 (19)	47 (15)	42 (25)	.46[b]
BMI (kg/m^2), n (%)				.96[a]
>30	11 (12)	5 (12)	6 (11)	
25-30	28 (29)	12 (29)	16 (30)	
<25	56 (59)	25 (60)	31 (58)	
Household size, median (IQR)	2 (2)	2 (2)	3 (2)	.61[b]
Recruiting source, n (%)				.50[a]
Helsana	76 (80)	33 (79)	43 (81)	
University	5 (5)	1 (2)	4 (8)	
Social media	6 (6)	3 (7)	3 (6)	
Other	8 (8)	5 (12)	3 (6)	

[a]Pearson chi-square test.

[b]Mann-Whitney *U* test.

At baseline, the median steps per day were 4624 (IQR 4497.79), that is, roughly half the number of steps per day compared with a representative Swiss sample [72]. The median nutritional quality of food purchases was 6.13 FSA-NPS DI points (IQR 4.02), which is in line with the values reported in a large-scale study [73]. Food purchases were low sugar (median 0.88 FSA-NPS DI points, IQR 1.20) and contained moderate amounts of saturated fatty acids (median 3.01 FSA-NPS DI points, IQR 1.66), sodium (median 2.50 FSA-NPS DI points, IQR 1.40), and protein (median 3.16 FSA-NPS DI points, IQR 0.93). The shares of fruits, vegetables, legumes, and nuts, and fiber were rather low (median 0.30 FSA-NPS DI points, IQR 0.48 and median 1.53 FSA-NPS DI points, IQR 0.77, respectively). No baseline differences were found between the intervention and control groups with regard to physical activity or the total nutritional quality of food purchases. Participants in the FutureMe intervention had significantly higher amounts of saturated fatty acids in their food purchases (*P*=.03) compared with the findings in the control group. Overall, participants showed high levels of motivational and recovery self-efficacy (median 6.0, IQR 1.25 and median 6.0, IQR 1.50, respectively), as well as outcome expectancy (median 6.2, IQR 1.60). Participants in the intervention group scored significantly higher on recovery self-efficacy than those in the control group (*P*=.03). Participants exhibited high intrinsic motivation to live a healthy lifestyle (median 6.0, IQR 1.33) and were less extrinsically motivated (median 3.0, IQR 2.33). Baseline characteristics with regard to all outcomes are presented in Table 3.

Table 3. Primary and secondary outcomes at baseline.

Primary and secondary outcomes	Total population		Intervention group		Control group		*P* value[a]
	Value, median (IQR)	n	Value, median (IQR)	n	Value, median (IQR)	n	
Physical activity							
Walking (steps/day)	4624.00 (4497.79)	92	4314.83 (3620.33)	41	5084.83 (5220.33)	51	.53
Nutritional quality of food purchases							
Total nutritional quality (FSA-NPS DI[b] points)[c]	6.13 (4.02)	67	6.73 (5.03)	27	5.83 (4.11)	40	.14
Nutritional subcategories							
Sugars (FSA-NPS DI points)[d]	0.88 (1.20)	67	0.83 (1.08)	27	0.91 (1.31)	40	.73
Saturated fatty acids (FSA-NPS DI points)[d]	3.01 (1.66)	67	3.82 (2.83)	27	2.93 (1.37)	40	.03
Sodium (FSA-NPS DI points)[d]	2.50 (1.40)	67	2.77 (1.56)	27	2.42 (1.37)	40	.42
Fruits, vegetables, legumes, and nuts (FSA-NPS DI points)[e]	0.30 (0.48)	67	0.28 (0.53)	27	0.31 (0.47)	40	.64
Fiber (FSA-NPS DI points)[e]	1.53 (0.77)	67	1.49 (0.70)	27	1.56 (0.83)	40	.85
Protein (FSA-NPS DI points)[e]	3.16 (0.93)	67	3.20 (0.70)	27	3.15(1.03)	40	.31
Attitudinal and motivational constructs							
Motivational self-efficacy[f]	6.00 (1.25)	95	6.00 (1.31)	42	6.00 (1.50)	53	.65
Recovery self-efficacy[f]	6.00 (1.50)	95	6.00 (1.50)	42	6.00 (1.50)	53	.03
Outcome expectancy[f]	6.20 (1.60)	95	6.10 (1.65)	42	6.20 (1.60)	53	.31
Intrinsic motivation[f]	6.00 (1.33)	95	5.83 (1.75)	42	6.00 (1.50)	53	.13
Extrinsic motivation[f]	3.00 (2.33)	95	2.67 (2.00)	42	3.33 (2.00)	53	.65

[a]Mann-Whitney *U* test.

[b]FSA-NPS DI: British Food Standards Agency Nutrient Profiling System Dietary Index.

[c]FSA-NPS DI point scale: −15 most healthy to +40 least healthy.

[d]FSA-NPS DI point scale: 0 most healthy to +10 least healthy.

[e]FSA-NPS DI point scale: 0 least healthy to +5 most healthy.

[f]Likert scale: 1 [completely disagree] to 7 [completely agree].

Physical Activity

As shown in Table 4, participants from both groups increased their physical activity levels within the first 4 weeks of the intervention (FutureMe T1: median 4577.85 steps/day, IQR 4620.26; control T1: median 5361.50 steps/day, IQR 5960.88). Participants in the intervention group consistently increased their physical activity over the course of the intervention from median 4314.81 (IQR 3620.33) steps/day at baseline to median 6042.31 (IQR 6242.31) steps/day at T2 and to median 4556.94 (IQR 5226.94) steps/day at the end of the intervention. In contrast, in the control group, the number of steps per day decreased after a short-term increase in T1, from median 5084.83 (IQR 5220.33) steps/day at baseline to median 2909.32 (IQR 5128.09) steps/day at T2 and to median 4821.64 (IQR 5228.79) steps/day at the end of the intervention. However, for both groups, the changes in the number of steps over the course of the intervention were not statistically significant. There were no significant differences in steps per day evident between the intervention and control groups.

Table 4. Primary outcomes by timepoint and group.

Primary outcome variables and groups	Baseline (T0)[a]		T1[b]		T2[c]		T3[d]	
	Value, median (IQR)	n	Value, median (IQR)	n	Value, median (IQR)	n	Value, median (IQR)	n
Physical activity, walking (steps/day)								
Avatar	4314.83 (3620.33)	41	4577.85 (4620.26)	35	6042.45 (6242.31)	18	4556.94 (5226.94)	12
Control	5084.83 (5220.33)	51	5361.50 (5960.88)	43	4909.32 (5128.09)	24	4821.64 (5228.79)	18
Nutritional quality of food purchases, total nutritional quality (FSA-NPS DI[e]points)[f]								
Avatar	6.73 (5.03)	27	6.79 (3.67)	20	6.14 (2.63)	14	5.45 (3.19)	13
Control	5.83 (4.11)	40	5.79 (5.41)	34	4.42 (5.42)	19	4.86 (3.75)	14

[a]Baseline: For steps, baseline is defined as average steps per day 6 days prior to enrolling in the trial, and for shopping-related outcome variables, baseline is defined as the nutritional value of all foods purchased within the 4 weeks before the trial.

[b]T1: week 1-4 average values.

[c]T2: week 5-8 average values.

[d]T3: week 9-12 average values (end of study).

[e]FSA-NPS DI: British Food Standards Agency Nutrient Profiling System Dietary Index.

[f]FSA-NPS DI point scale: −15 most healthy to +40 least healthy.

Nutritional Quality of Food Purchases

Participants in both groups improved the nutritional quality of their food purchases over the course of the intervention (Multimedia Appendix 4), albeit insignificantly. Participants in the intervention group shopped less healthily within the first 4 weeks of the intervention (median 6.79 FSA-NPS DI points, IQR 3.67) compared with the findings at baseline (median 6.73 FSA-NPS DI points, IQR 5.03). However, they improved their shopping behavior in T2 (median 6.14 FSA-NPS DI points, IQR 2.63) and at the end of the intervention (median 5.45 FSA-NPS DI points, IQR 3.19). Participants in the control group consistently improved their shopping behavior over the course of the intervention from median 5.83 FSA-NPS DI points (IQR 4.11) at baseline to median 5.70 FSA-NPS DI points (IQR 5.31) at T1, median 4.42 FSA-NPS DI points (IQR 5.42) at T3, and median 4.86 FSA-NPS DI points (IQR 3.75) at the end of the intervention. At the end of the intervention, the control group purchased significantly healthier food than the FutureMe group (P=.02). However, the control group's food baskets were already healthier at baseline, although not at statistically significant levels.

Nutritional Subcategories

Multimedia Appendix 5 provides a summary of all secondary outcomes for all relevant timepoints. It also presents the results of the Wilcoxon signed-rank tests and Mann-Whitney U tests. Across both groups, no statistically significant improvements were found in nutritional subcategories during the intervention. As for directional developments, for the control group, improvements in the nutritional quality of food purchases were driven by reductions in sugar per 100 g food purchases between baseline (median 0.91 FSA-NPS DI points, IQR 1.31) and the end of the study (median 0.55 FSA-NPS DI points, IQR 1.07). For the intervention group, improvements in the nutritional quality of food purchases were driven by reductions in saturated fatty acids and sodium between baseline (saturated fatty acids: median 3.82 FSA-NPS DI points, IQR 2.83; sodium: median 2.77 FSA-NPS DI points, IQR 1.56) and the end of the study (saturated fatty acids: median 3.82 FSA-NPS DI points, IQR 2.83; sodium: median 2.77 FSA-NPS DI points, IQR 1.56). In both groups, the amounts of protein, fruits, vegetables, and fiber per 100 g in food purchases were relatively stable throughout the trial or even decreased slightly. At the end of the intervention, the control group purchased significantly healthier food than the FutureMe group with regard to sugar (P=.01) and saturated fatty acids (P=.01). However, the control group's food baskets already contained significantly lower amounts of saturated fatty acids at baseline (P=.03).

User Engagement

User engagement was similar across the FutureMe and control interventions (see Multimedia Appendix 5). Engagement was the highest during the first 4 weeks of the intervention (T1), where the FutureMe and control groups recorded a median of 5 logins per week (IQR 10) and 4 logins per week (IQR 10), respectively. User engagement decreased to 1 login per week (IQR 3) for the FutureMe group for T2 and remained flat (IQR 1) until the end of the study. For the control group, user engagement also dropped to a median of 1 login per week (IQR 3) at T2 but increased to 2 logins per week (IQR 2) at the end of the study.

Attitudinal and Motivational Constructs

Multimedia Appendix 5 provides a summary of all attitudinal and motivational constructs preintervention and postintervention. It also presents the results of the Wilcoxon signed-rank tests and the Mann-Whitney U tests. Intrinsic motivation to lead a healthy lifestyle significantly (P=.03) increased from a median value of 5.83 (IQR 1.75) at baseline to 6.0 (IQR 1.67) at the end of the study for participants in the FutureMe group. In contrast, intrinsic motivation decreased throughout the intervention in the control group. The extrinsic motivation to lead a healthy lifestyle increased in both groups from baseline (FutureMe: median 2.67, IQR 2.00; control: median 3.33, IQR 2.00) to the end of the study (FutureMe: median 3.67, IQR 2.50;

control: median 4.00, IQR 2.08). However, the changes were not statistically significant. At the end of the study, the FutureMe group showed significantly (*P*=.03) higher levels of recovery self-efficacy (median 7.00, IQR 1.00) compared with the levels in the control group (median 5.67, IQR 1.25). Motivational self-efficacy did not change significantly in either group. Outcome expectancy increased in the FutureMe group between baseline (median 6.10, IQR 1.65) and the end of the study (median 7.00, IQR 2.08). However, the change was not statistically significant. In contrast, outcome expectancy decreased in the control group between baseline (median 6.20, IQR 1.60) and the end of the study (median 5.67, IQR 1.62), although not at statistically significant levels.

Sensitivity Analysis

To test for attrition bias and randomness of missing data, we compared demographic and primary outcome variables at baseline (T0) between participants who did not finish the trial (ie, dropouts) and participants who completed the trial (ie, trial completers). The results are presented in Table 5. We found no statistical differences between dropouts and trial completers with respect to our primary outcome variables, indicating a low risk of attrition bias. Comparing demographic variables revealed no statistical differences with regard to gender ratio, BMI, or household size. Trial completers, however, were found to be significantly younger (*P*=.02) than dropouts. Overall, our sensitivity analysis confirmed that missing data can be considered completely random in our trial and that no systematic error can thus be expected [70]. The large amount of missing data, however, reduced the statistical power of our analysis [69,70].

Table 5. Results of the sensitivity analysis.

Demographics and primary outcomes	Dropouts		Trial completers		*P* value[a]
	Value	n	Value	n	
Gender, n (%)		65		30	.66
Male	28 (43)		15 (50)		
Female	37 (57)		15 (50)		
Age (years), median (IQR)	48 (20)	65	42 (18)	30	.02
BMI (kg/m^2), n (%)		65		30	.92[b]
>30	9 (14)		1 (3)		
25-30	18 (28)		11 (37)		
<25	38 (58)		18 (60)		
Household size, median (IQR)	2 (2)		3 (2)		.54
Physical activity, median (IQR)					
Walking (steps/day)	4897.83 (4286.13)	64	4077.50 (5499.05)	28	.95
Nutritional quality of food purchases, median (IQR)					
Total nutritional quality (FSA-NPS DI[c] points)[d]	5.64 (3.87)	42	6.38 (4.26)	28	.24

[a]Mann-Whitney *U* test.

[b]Mann-Whitney *U* test was performed comparing absolute BMI scores.

[c]FSA-NPS DI: British Food Standards Agency Nutrient Profiling System Dietary Index.

[d]FSA-NPS DI point scale: −15 most healthy to +40 least healthy.

Qualitative Study Results

Among all participants who finished the trial, 15 participants were randomly selected to participate in semistructured video interviews. The selected participant demographics were similar to the overall trial demographics, and 47% (7/15) of the participants were female. The median age was 34 years (IQR 24), and the median household size was 2 (IQR 1). Overall, 53% (8/15) of the participants had received the FutureMe intervention and 47% (7/15) had received the control intervention.

Ease of Use and Usefulness

While most participants experienced some challenges while signing up, particularly with connecting their loyalty cards to the app, they found using the app intuitive and easy. App functionalities were said to be simple and easily understandable.

> *The app is very simple, easy to read and intuitive to use. Really without a lot of bells and whistles or hidden features.* [User #13, female]

Overall, many users appreciated how easy the app was to use after the initial setup and that data synchronized automatically and effortlessly into the app.

> *Personally, I was motivated by the fact that the effort is relatively low, it is automatically synchronized, whether it is sports or shopping.* [User #1, male]

The app's observed usefulness depended largely on how well participants perceived that tracking reflected their actual physical activity and food purchasing behavior. Participants who shopped mainly at the 2 participating grocery retailers and whose physical activity behavior mostly involved walking or running, rated app usefulness as very high. They felt that the app provided helpful insights, liked the color guidance on the overview tracking screen, and were inspired by the food tips.

> *Above all, to bring the bar up to green. As far as possible everywhere, but I haven't achieved that yet.* [User #8, male]

> *I just found it very exciting with the purchases. […] I had the feeling that I buy very consciously and healthily, and sometimes I was in the orange and red area, where I thought just because I bought a pizza once, it drags the whole curve down. That is then nevertheless exciting and stimulates you to think about it and to consider whether it would not make sense to buy the wholewheat pizza dough or a vegan alternative instead of the normal pizza.* [User #9, male]

Participants who purchased their food at different grocery retailers or local markets, who regularly ate takeaways, or who shared the loyalty card with other family members found the app less useful. They felt that it did not reflect their actual behavior. They also stated that they would rate usefulness higher if tracking were more accurate.

> *It does not reflect things as they are. Because I don't buy a lot of vegetables and fruit at these two retailers.* [User #14, male]

> *I was annoyed that my son just once again bought chips and peanuts.* [User #5, female]

Participant-Avatar Relationship

Users who rated the app's tracking accuracy high reported that the avatar motivated behavior change and that checking-in on the avatar was a key reason for logging into the app.

> *The avatar was super cool.* [User #6, male]

> *One app feature that motivated me was the avatar. I always clicked on the different items and checked to make sure that I improve on all categories.* [User #4, female]

Users liked that they could personalize their avatar and mentioned that the avatar somewhat resembled them. In contrast, most participants did not much like the avatar's *mood* feature (ie, the facial expression depending on fruit, vegetable, and fiber consumption). In particular, when the avatar expressed sadness, users reported that this did not reflect how they felt. Also, more generally, users with less healthy food purchases or low physical activity behavior were more critical, with some users feeling discouraged by the avatar's challenging feedback. They felt that some of the avatar's reactions to their behaviors were exaggerated and did not reflect their state of health. This suggests that users struggled to fully understand that their avatar did not reflect their current health but rather their future health.

> *I could only partially identify with the avatar. I found especially that he gained weight very quickly. I couldn't understand why it was so fast. I had the feeling that he represented a worse image than my actual condition.* [User #11, male]

Users who rated the app's tracking accuracy as low identified less with the avatar.

> *I mostly buy drinks and non-food items at the two participating retailers, so nothing that affects the avatar. Therefore, I can't identify with it.* [User #15, female]

Areas of Improvement

Users suggested improvements to the app's tracking functionality; its analysis, insights, and feedback; and incentive schemes.

The most frequently mentioned area of improvement concerned broadening synchronized data sources and improving tracking accuracy. Although users appreciated that the automatic synchronization of food purchases required no personal effort, they wanted to be able to manually add data from other retailers, to include restaurant consumption, or to exclude foods purchased by other household members or unconsumed (food waste). Receipt scanning was mentioned multiple times as a preferred option for manually adding data, followed by drop-down functionalities. Further, users indicated that breaking down food purchases over a consumption period would improve app usefulness as certain items (eg, oil, salt, or condiments) are usually consumed over a longer period of time.

> *It would be cool if you could actually control the food analysis in the app via consumption and not the purchase, by dividing a product according to days or months in which I consume it.* [User #1, male]

Regarding activity tracking, users felt that including physical activity beyond the number of steps and improving app compatibility with leading fitness watches would be valuable improvements. Another related improvement often mentioned was improving data synchronization speed. In the FutureMe trial, food and physical activity data were updated daily. However, as an app login was required to synchronize loyalty card data with the app, it took approximately 2 days for food purchases to appear in the app. Users mentioned that they would have preferred to have food purchases available immediately after shopping.

Regarding the app's analysis options, users liked the food tips, but felt that these were not updated regularly enough. Some users indicated that they would have liked to receive push notifications once new data were available in the app, or motivating push messages if their shopping or physical activity behavior deteriorated or improved. Furthermore, various participants mentioned that they would have liked to see the least healthy items in their shopping basket on a separate screen, as they felt that this would help to improve their shopping

behavior, instead of merely receiving shopping tips for healthier alternatives.

> *You only see alternative product suggestions in the app. But it would be cool to see which products you purchased that you should consume less of. That would be more helpful.* [User #4, female]

Regarding feedback, some participants were discouraged by the avatar's sometimes harsh language or appearance. They felt that empathetic positive language, even if they shopped unhealthily, would be more encouraging, which aligns with the motivational interviewing theory [74]. Furthermore, participants mentioned that animating the avatar or being able to adjust its clothing during the trial would improve user engagement with the overall app.

> *If you could dress the avatar however you want, I'd love that.* [User #15, female]

Lastly, some users suggested implementing an extrinsic incentive scheme in the app. Various users mentioned that an integrated financial bonus system or discounts for healthier food options would increase their motivation to use the app.

Discussion

Principal Findings

Insufficient physical activity and unhealthy diets are contributing to the rise in NCDs. Scalable preventative interventions are needed to combat this trend. To date, however, evidence of scalable mHealth interventions targeting both physical activity and nutritional improvements is scarce.

This is the first study to examine the impact of a future-self avatar mHealth intervention on physical activity and on the nutritional quality of food purchases. Using a 12-week RCT design, we found small statistically insignificant improvements in physical activity (median +242 steps/day) and small insignificant improvements in the nutritional quality of food purchases (median −1.28 FSA-NPS DI points). Low nutritional quality of food purchases is associated with a higher risk of developing chronic diseases and higher levels of obesity [60]. Moreover, increases in physical activity can lower the risk of metabolic syndromes and reduce all-cause mortality [75,76]. Our results provide some first evidence that scalable preventative mHealth interventions that use future-self avatars and automatic tracking of food choices via digital receipts from loyalty cards could contribute to reducing NCDs.

However, our study found no statistically significant changes in either of its primary outcomes over the course of the 12-week intervention. One underlying reason why we found no statistically significant improvements might be the reduced statistical power of our research due to the unexpectedly high attrition rate (n=95, 68.4%) and the resulting small sample at the end of the intervention. A meta-review of weight-loss interventions found that attrition rates vary significantly between studies (9%-90%) but found no consistent predictors [77]. Also, prior research has noted that high dropout rates are a "natural feature" of mHealth interventions [20]. Compared with traditional randomized controlled drug trials, adherence to an mHealth intervention is largely at the participant's discretion [20]. This applies in particular when an intervention is not critical to participant well-being, as was the case in our preventative mHealth study. Further research with larger sample sizes is thus needed to understand whether a future-self avatar mHealth intervention can effectively reduce present bias and drive health behavior change.

To our knowledge, our study is the first mHealth intervention to implement fully automated food purchase tracking through loyalty cards in an RCT design outside of South Africa [25,26]. Based on a 12-week RCT and 15 semistructured interviews, we found that our technical solution is feasible and reliably tracks food purchases without requiring any participant effort after the initial setup of the FutureMe app. Prior work has found that individuals from northern European countries consume between 71.5% and 79.8% of their mean daily energy intake at home [78]. Analyzing food purchases through automized receipt tracking and a comprehensive nutritional product database is thus a promising method to guide improvements in overall nutrition and contributes to the need for automated and scalable nutrition tracking solutions [25,26]. Our study found that some participants were open to further improving the comprehensiveness of the analyzed food data by manually adding out-of-home food consumption and by increasing the number of included retailers. Future research should thus evaluate how to further optimize automated and semiautomated nutrition tracking systems so as to increase user retention.

Present bias offers a possible explanation as to why preventative mHealth interventions are less effective than curative mHealth interventions [15,32]. Our study thus examined whether a future-self avatar intervention can reduce present bias by increasing outcome expectancy. In our 12-week RCT, we found small statistically insignificant improvements in outcome expectancy in the FutureMe group, while this decreased in the control group. The results of our qualitative interviews show that some users did not understand that the FutureMe avatar reflected the future consequences of their current health behaviors. They assumed instead that the FutureMe avatar mirrors today's health consequences. This effect could have undermined significant increases in outcome expectancy. However, we found that the future-self avatar intervention significantly increased intrinsic motivation (P=.03), while no increase occurred with the control intervention. Intrinsic motivation is associated with more sustainable behavior change than extrinsically motivated behavior [79]. Future-self avatars, as an element of gamification, might thus meaningfully contribute to improving the longer-term effectiveness of mHealth interventions, compared to more traditional text-based and numerical interfaces. Future research should explore how the interfaces and functionalities of future-self avatar mHealth interventions might be improved to more clearly communicate future health consequences, given the promising results this technology has produced in overcoming present bias in related research fields [31].

Limitations

The strengths of our study are its use of objectively measured data for all primary and secondary outcomes, its rigorous

research design, and its innovative future-self avatar interface and fully automated tracking system. However, it is not without limitations. First, the unexpectedly high attrition rate limited the sample size and statistical power of our study. Our qualitative study revealed that higher than expected dropout was partly caused by some users struggling to connect their grocery loyalty cards to the FutureMe app. Furthermore, users who shared the loyalty card with other members of their household or who did not shop at participating retailers for significant parts of their diets found the app less useful, potentially causing further dropout. We conducted a sensitivity analysis to test for attrition bias and found that participants who discontinued the app were not different from participants who finished the trial, based on their baseline demographics (with the exception of age), physical activity, and food purchasing behavior.

Second, the timing of our study may have impacted our results. Prior research has found that physical activity behavior is seasonally dependent [80]. Our study was conducted from November 2020 to April 2021, and the results thus may have been influenced by decreases in daylight and temperatures, as well as by unusual physical activity or food purchasing behavior due to the end-of-year holiday season and New Year's resolutions. Also, the rise of the COVID-19 pandemic during the trial period may have impacted physical activity patterns and grocery purchases.

Third, we only tracked and analyzed food purchased at Switzerland's 2 leading grocery retailers. The qualitative study demonstrated that this methodology does not comprehensively capture nutrition for certain user groups. However, the chosen retailers represent 69% of the grocery market in Switzerland [81]. While this share is considerable, noninclusion of other food purchases (eg, local farmers' markets, butcher stores, or discounters) may have impacted our results. Also, data on purchased groceries automatically collected via loyalty cards usually only represent a subset of consumed products. In fact, the data set is missing information about food items sourced at other stores or markets, meals consumed in restaurants, and which of the purchased items were not consumed by the respective app user (eg, food waste or products eaten by household members other than the buyer). Furthermore, physical activity tracking was limited to measuring steps and did not include other forms of physical activity, which may have constrained our findings. Additionally, for iOS users, we only captured the number of steps recorded by their mobile phones. For Android users, we captured both the steps recorded by their mobile phones and those recorded by other devices connected to Google Fit.

Finally, the context of human diets is complex as individual dietary needs depend on many factors often unknown to an automatic system, including lifestyle, food allergies, and diet-related diseases, rendering static interventions ineffective. In the absence of a dietician or physician who could mitigate these challenges, automatic assessment interventions must be carefully designed, based on established nutritional guidelines (such as the Nutri-Score framework), and limited to healthy population samples.

Conclusions

Our RCT found that a future-self avatar mHealth intervention significantly increased intrinsic motivation to lead a healthy lifestyle but did not lead to statistically significant improvements in physical activity levels and food purchasing behavior. Leveraging loyalty card data to track the nutritional quality of food purchases was found to be a feasible and accepted nutrition tracking technology and thus contributes to the search for scalable and automated tracking solutions. However, the high attrition rate and resulting small sample size of our study require further research to evaluate whether future-self avatar mHealth interventions can improve physical activity and food purchases at statistically significant levels. The use of future-self avatars may need to be modified in order to prevent attrition, reduce present bias, and promote sustainable behavior change.

Acknowledgments

We thank the Swiss National Science Foundation (project #188402) and Helsana Zusatzversicherungen AG for financially supporting this study. Helsana supported participant recruitment but was not involved in the study design, data analysis, and interpretation, or in reviewing and approving the manuscript for publication. Michael Zeltner developed the front-end app. BitsaboutMe supported the back-end integration of loyalty card data.

Authors' Contributions

AM, KF, and SM conceived and designed the study. KF oversaw the development of the nutritional product database. AM led the research project, wrote the trial protocol, performed the data analysis, and wrote the manuscript. JW performed the food purchase data analysis. JA provided technical support for data updates. All authors provided critical reviews on earlier drafts of the manuscript and approved the final version of the manuscript before submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary information on grocery loyalty card programs in Switzerland.
[PDF File (Adobe PDF File), 183 KB - jmir_v24i7e32487_app1.pdf]

Multimedia Appendix 2
Calculation scheme for FutureMe avatar health states.
[PDF File (Adobe PDF File), 193 KB - jmir_v24i7e32487_app2.pdf]

Multimedia Appendix 3
Scales for attitudinal and motivational constructs used in the FutureMe randomized controlled trial.
[PDF File (Adobe PDF File), 376 KB - jmir_v24i7e32487_app3.pdf]

Multimedia Appendix 4
Primary outcomes by timepoint, within- and between-group comparisons.
[PDF File (Adobe PDF File), 48 KB - jmir_v24i7e32487_app4.pdf]

Multimedia Appendix 5
Secondary outcomes by timepoint and group, within- and between-group comparisons.
[PDF File (Adobe PDF File), 258 KB - jmir_v24i7e32487_app5.pdf]

Multimedia Appendix 6
CONSORT-eHEALTH checklist (V 1.6.2).
[PDF File (Adobe PDF File), 92 KB - jmir_v24i7e32487_app6.pdf]

References

1. Benziger CP, Roth GA, Moran AE. The Global Burden of Disease Study and the preventable burden of NCD. Glob Heart 2016 Dec;11(4):393-397. [doi: 10.1016/j.gheart.2016.10.024] [Medline: 27938824]
2. Herausforderung nichtübertragbare Krankheiten. Bundesamt für Gesundheit. URL: https://www.bag.admin.ch/dam/bag/de/dokumente/nat-gesundheitsstrategien/ncd-strategie/ncd-strategie_in_kuerze.pdf.download.pdf/NCD-Strategie_in_kuerze.pdf [accessed 2022-05-30]
3. Blakely T, Kvizhinadze G, Atkinson J, Dieleman J, Clarke P. Health system costs for individual and comorbid noncommunicable diseases: An analysis of publicly funded health events from New Zealand. PLoS Med 2019 Jan 8;16(1):e1002716 [FREE Full text] [doi: 10.1371/journal.pmed.1002716] [Medline: 30620729]
4. Global Action Plan for the Prevention and Control of NCDs 2013-2020. World Health Organization. 2013. URL: https://www.who.int/publications/i/item/9789241506236 [accessed 2021-06-01]
5. Aburto NJ, Ziolkovska A, Hooper L, Elliott P, Cappuccio FP, Meerpohl JJ. Effect of lower sodium intake on health: systematic review and meta-analyses. BMJ 2013 Apr 03;346:f1326 [FREE Full text] [doi: 10.1136/bmj.f1326] [Medline: 23558163]
6. Mozaffarian D, Fahimi S, Singh GM, Micha R, Khatibzadeh S, Engell RE, et al. Global sodium consumption and death from cardiovascular causes. N Engl J Med 2014 Aug 14;371(7):624-634. [doi: 10.1056/nejmoa1304127]
7. Afshin A, Sur PJ, Fay KA, Cornaby L, Ferrara G, Salama JS, et al. Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet 2019 May;393(10184):1958-1972. [doi: 10.1016/S0140-6736(19)30041-8]
8. Lee I, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. The Lancet 2012 Jul;380(9838):219-229. [doi: 10.1016/S0140-6736(12)61031-9]
9. Sallis JF, Bull F, Guthold R, Heath GW, Inoue S, Kelly P, et al. Progress in physical activity over the Olympic quadrennium. The Lancet 2016 Sep;388(10051):1325-1336. [doi: 10.1016/S0140-6736(16)30581-5]
10. Bianchini F, Kaaks R, Vainio H. Weight control and physical activity in cancer prevention. Obes Rev 2002 Mar;3(1):5-8. [doi: 10.1046/j.1467-789x.2002.00046.x] [Medline: 12119660]
11. Choi KW, Chen C, Stein MB, Klimentidis YC, Wang M, Koenen KC, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. Assessment of bidirectional relationships between physical activity and depression among adults: A 2-sample mendelian randomization study. JAMA Psychiatry 2019 Apr 01;76(4):399-408 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.4175] [Medline: 30673066]
12. Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. The Lancet Global Health 2018 Oct;6(10):e1077-e1086. [doi: 10.1016/s2214-109x(18)30357-7]
13. Powles J, Fahimi S, Micha R, Khatibzadeh S, Shi P, Ezzati M, Global Burden of Diseases NutritionChronic Diseases Expert Group (NutriCoDE). Global, regional and national sodium intakes in 1990 and 2010: a systematic analysis of 24 h urinary sodium excretion and dietary surveys worldwide. BMJ Open 2013 Dec 23;3(12):e003733 [FREE Full text] [doi: 10.1136/bmjopen-2013-003733] [Medline: 24366578]

14. Thout SR, Santos JA, McKenzie B, Trieu K, Johnson C, McLean R, et al. The Science of Salt: Updating the evidence on global estimates of salt intake. J Clin Hypertens (Greenwich) 2019 Jun 29;21(6):710-721 [FREE Full text] [doi: 10.1111/jch.13546] [Medline: 31033166]
15. Mönninghoff A, Kramer JN, Hess AJ, Ismailova K, Teepe GW, Tudor Car L, et al. Long-term effectiveness of mHealth physical activity interventions: Systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2021 Apr 30;23(4):e26699 [FREE Full text] [doi: 10.2196/26699] [Medline: 33811021]
16. Gal R, May AM, van Overmeeren EJ, Simons M, Monninkhof EM. The effect of physical activity interventions comprising wearables and smartphone applications on physical activity: a systematic review and meta-analysis. Sports Med Open 2018 Sep 03;4(1):42 [FREE Full text] [doi: 10.1186/s40798-018-0157-9] [Medline: 30178072]
17. 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]
18. Smith D, Duque L, Huffman J, Healy B, Celano C. Text message interventions for physical activity: A systematic review and meta-analysis. Am J Prev Med 2020 Jan;58(1):142-151 [FREE Full text] [doi: 10.1016/j.amepre.2019.08.014] [Medline: 31759805]
19. Ali SH, Luo R, Li Y, Liu X, Tang C, Zhang P. Application of mobile health technologies aimed at salt reduction: Systematic review. JMIR Mhealth Uhealth 2019 Apr 17;7(4):e13250. [doi: 10.2196/13250] [Medline: 30994467]
20. 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]
21. Brindal E, Freyne J, Saunders I, Berkovsky S, Smith G, Noakes M. Features predicting weight loss in overweight or obese participants in a web-based intervention: randomized trial. J Med Internet Res 2012 Dec 12;14(6):e173 [FREE Full text] [doi: 10.2196/jmir.2156] [Medline: 23234759]
22. Laing BY, Mangione CM, Tseng C, Leng M, Vaisberg E, Mahida M, et al. Effectiveness of a smartphone application for weight loss compared with usual care in overweight primary care patients: a randomized, controlled trial. Ann Intern Med 2014 Nov 18;161(10 Suppl):S5-12 [FREE Full text] [doi: 10.7326/M13-3005] [Medline: 25402403]
23. Fuchs K, Haldimann M, Vuckovac D, Ilic A. Automation of Data Collection Techniques for Recording Food Intake: a Review of Publicly Available and Well-Adopted Diet Apps. 2018 Presented at: 2018 International Conference on Information and Communication Technology Convergence (ICTC); October 17-19, 2018; Jeju, Korea (South) p. 58-65. [doi: 10.1109/ictc.2018.8539468]
24. Svensson Å, Larsson C. A mobile phone app for dietary intake assessment in adolescents: An evaluation study. JMIR Mhealth Uhealth 2015 Nov 03;3(4):e93 [FREE Full text] [doi: 10.2196/mhealth.4804] [Medline: 26534783]
25. Sturm R, An R, Segal D, Patel D. A cash-back rebate program for healthy food purchases in South Africa: results from scanner data. Am J Prev Med 2013 Jun;44(6):567-572 [FREE Full text] [doi: 10.1016/j.amepre.2013.02.011] [Medline: 23683973]
26. An R, Patel D, Segal D, Sturm R. Eating better for less: A national discount program for healthy food purchases in South Africa. Am J Hlth Behav 2013 Jan 01;37(1):56-61. [doi: 10.5993/ajhb.37.1.6]
27. Hassannejad H, Matrella G, Ciampolini P, De Munari I, Mordonini M, Cagnoni S. Automatic diet monitoring: a review of computer vision and wearable sensor-based methods. Int J Food Sci Nutr 2017 Sep 31;68(6):656-670. [doi: 10.1080/09637486.2017.1283683] [Medline: 28139173]
28. Schwarzer R. Modeling health behavior change: How to predict and modify the adoption and maintenance of health behaviors. Applied Psychology 2008 Jan;57(1):1-29. [doi: 10.1111/j.1464-0597.2007.00325.x]
29. Wang Y, Sloan FA. Present bias and health. J Risk Uncertain 2018 Oct;57(2):177-198 [FREE Full text] [doi: 10.1007/s11166-018-9289-z] [Medline: 31244508]
30. Ersner-Hershfield H, Garton MT, Ballard K, Samanez-Larkin GR, Knutson B. Don't stop thinking about tomorrow: Individual differences in future self-continuity account for saving. Judgm Decis Mak 2009 Jun 01;4(4):280-286 [FREE Full text] [Medline: 19774230]
31. Hershfield HE, Goldstein DG, Sharpe WF, Fox J, Yeykelis L, Carstensen LL, et al. Increasing saving behavior through age-progressed renderings of the future self. Journal of Marketing Research 2019 Mar 06;48(SPL):S23-S37. [doi: 10.1509/jmkr.48.spl.s23]
32. Hunter RF, Tang J, Hutchinson G, Chilton S, Holmes D, Kee F. Association between time preference, present-bias and physical activity: implications for designing behavior change interventions. BMC Public Health 2018 Dec 19;18(1):1388 [FREE Full text] [doi: 10.1186/s12889-018-6305-9] [Medline: 30567532]
33. Baicker K, Mullainathan S, Schwartzstein J. Behavioral Hazard in health insurance. Q J Econ 2015 Nov;130(4):1623-1667 [FREE Full text] [doi: 10.1093/qje/qjv029] [Medline: 35602854]
34. Mitchell MS, Goodman JM, Alter DA, John LK, Oh PI, Pakosh MT, et al. Financial incentives for exercise adherence in adults: systematic review and meta-analysis. Am J Prev Med 2013 Nov;45(5):658-667. [doi: 10.1016/j.amepre.2013.06.017] [Medline: 24139781]
35. Kramer J, Tinschert P, Scholz U, Fleisch E, Kowatsch T. A cluster-randomized trial on small incentives to promote physical activity. Am J Prev Med 2019 Feb;56(2):e45-e54 [FREE Full text] [doi: 10.1016/j.amepre.2018.09.018] [Medline: 30553693]

36. Wall J, Mhurchu CN, Blakely T, Rodgers A, Wilton J. Effectiveness of monetary incentives in modifying dietary behavior:a review of randomized, controlled trials. Nutr Rev 2006 Dec;64(12):518-531. [doi: 10.1111/j.1753-4887.2006.tb00185.x] [Medline: 17274494]
37. Etkin J. The hidden cost of personal quantification. J Consum Res 2016 Feb 16;42(6):967-984. [doi: 10.1093/jcr/ucv095]
38. Kruglanski AW, Riter A, Amitai A, Margolin B, Shabtai L, Zaksh D. Can money enhance intrinsic motivation? A test of the content-consequence hypothesis. Journal of Personality and Social Psychology 1975 Apr;31(4):744-750. [doi: 10.1037/0022-3514.31.4.744]
39. Lockwood BB. Optimal income taxation with present bias. American Economic Journal: Economic Policy 2020 Nov 01;12(4):298-327. [doi: 10.1257/pol.20180762]
40. Navarro J, Cebolla A, Llorens R, Borrego A, Baños RM. Manipulating self-avatar body dimensions in virtual worlds to complement an internet-delivered intervention to increase physical activity in overweight women. Int J Environ Res Public Health 2020 Jun 05;17(11):4045 [FREE Full text] [doi: 10.3390/ijerph17114045] [Medline: 32517117]
41. Karekla M, Savvides S, Gloster A. An avatar-led intervention promotes smoking cessation in young adults: A pilot randomized clinical trial. Ann Behav Med 2020 Oct 01;54(10):747-760 [FREE Full text] [doi: 10.1093/abm/kaaa013] [Medline: 32383736]
42. Fuchs K, Meusburger D, Haldimann M, Ilic A. NutritionAvatar: designing a future-self avatar for promotion of balanced, low-sodium diet intention: framework design and user study. In: CHItaly '19: Proceedings of the 13th Biannual Conference of the Italian SIGCHI Chapter: Designing the next interaction. 2019 Presented at: 13th Biannual Conference of the Italian SIGCHI Chapter: Designing the next interaction; September 23-25, 2019; Padova, Italy p. 1-10. [doi: 10.1145/3351995.3352044]
43. Pinto MD, Hickman RL, Clochesy J, Buchner M. Avatar-based depression self-management technology: promising approach to improve depressive symptoms among young adults. Appl Nurs Res 2013 Mar;26(1):45-48 [FREE Full text] [doi: 10.1016/j.apnr.2012.08.003] [Medline: 23265918]
44. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMC Med 2010 Mar 24;8:18 [FREE Full text] [doi: 10.1186/1741-7015-8-18] [Medline: 20334633]
45. Tudor-Locke C, Hatano Y, Pangrazi RP, Kang M. Revisiting "how many steps are enough?". Med Sci Sports Exerc 2008 Jul;40(7 Suppl):S537-S543. [doi: 10.1249/MSS.0b013e31817c7133] [Medline: 18562971]
46. Tudor-Locke CE, Myers AM, Bell RC, Harris SB, Wilson Rodger N. Preliminary outcome evaluation of the First Step Program: a daily physical activity intervention for individuals with type 2 diabetes. Patient Education and Counseling 2002 May;47(1):23-28. [doi: 10.1016/s0738-3991(01)00169-0]
47. Sugiura H, Sugiura H, Kajima K, Mirbod SM, Iwata H, Matsuoka T. Effects of long-term moderate exercise and increase in number of daily steps on serum lipids in women: randomised controlled trial [ISRCTN21921919]. BMC Womens Health 2002 Jan 21;2(1):3 [FREE Full text] [doi: 10.1186/1472-6874-2-3] [Medline: 11846892]
48. Julia C, Fialon M, Galan P, Deschasaux-Tanguy M, Andreeva VA, Kesse-Guyot E, et al. Are foods 'healthy' or 'healthier'? Front-of-pack labelling and the concept of healthiness applied to foods. Br J Nutr 2021 May 05;127(6):948-952. [doi: 10.1017/s0007114521001458]
49. Julia C, Méjean C, Touvier M, Péneau S, Lassale C, Ducrot P, et al. Validation of the FSA nutrient profiling system dietary index in French adults-findings from SUVIMAX study. Eur J Nutr 2016 Aug 21;55(5):1901-1910. [doi: 10.1007/s00394-015-1006-y] [Medline: 26293977]
50. Wu J, Fuchs K, Lian J, Haldimann ML, Schneider T, Mayer S, et al. Estimating dietary intake from grocery shopping data-A comparative validation of relevant indicators in Switzerland. Nutrients 2021 Dec 29;14(1):159 [FREE Full text] [doi: 10.3390/nu14010159] [Medline: 35011033]
51. Egnell M, Ducrot P, Touvier M, Allès B, Hercberg S, Kesse-Guyot E, et al. Objective understanding of Nutri-Score Front-Of-Package nutrition label according to individual characteristics of subjects: Comparisons with other format labels. PLoS One 2018 Aug 23;13(8):e0202095 [FREE Full text] [doi: 10.1371/journal.pone.0202095] [Medline: 30138359]
52. Ahn S, Fox J, Hahm J. Using virtual doppelgängers to increase personal relevance of health risk communication. In: Bickmore T, Marsella S, Sidner C, editors. Intelligent Virtual Agents. IVA 2014. Lecture Notes in Computer Science, vol 8637. Cham: Springer; 2014:1-12.
53. Moore RK. A Bayesian explanation of the 'Uncanny Valley' effect and related psychological phenomena. Sci Rep 2012 Nov 16;2(1):864 [FREE Full text] [doi: 10.1038/srep00864] [Medline: 23162690]
54. Cerin E, Cain KL, Oyeyemi AL, Owen N, Conway TL, Cochrane T, et al. Correlates of agreement between accelerometry and self-reported physical activity. Med Sci Sports Exerc 2016 Jun;48(6):1075-1084 [FREE Full text] [doi: 10.1249/MSS.0000000000000870] [Medline: 26784274]
55. Duncan MJ, Wunderlich K, Zhao Y, Faulkner G. Walk this way: validity evidence of iphone health application step count in laboratory and free-living conditions. J Sports Sci 2018 Aug 28;36(15):1695-1704. [doi: 10.1080/02640414.2017.1409855] [Medline: 29179653]
56. Nolan M, Mitchell J, Doyle-Baker P. Validity of the Apple iPhone® /iPod Touch® as an accelerometer-based physical activity monitor: a proof-of-concept study. J Phys Act Health 2014 May;11(4):759-769. [doi: 10.1123/jpah.2011-0336] [Medline: 23575387]

57. Major MJ, Alford M. Validity of the iPhone M7 motion co-processor as a pedometer for able-bodied ambulation. J Sports Sci 2016 Dec 30;34(23):2160-2164. [doi: 10.1080/02640414.2016.1189086] [Medline: 27240005]
58. Julia C, Péneau S, Buscail C, Gonzalez R, Touvier M, Hercberg S, et al. Perception of different formats of front-of-pack nutrition labels according to sociodemographic, lifestyle and dietary factors in a French population: cross-sectional study among the NutriNet-Santé cohort participants. BMJ Open 2017 Jun 15;7(6):e016108 [FREE Full text] [doi: 10.1136/bmjopen-2017-016108] [Medline: 28619781]
59. Julia C, Touvier M, Méjean C, Ducrot P, Péneau S, Hercberg S, et al. Development and validation of an individual dietary index based on the British Food Standard Agency nutrient profiling system in a French context. J Nutr 2014 Dec;144(12):2009-2017. [doi: 10.3945/jn.114.199679] [Medline: 25411035]
60. Julia C, Hercberg S, World Health Organization. Regional Office for Europe. Development of a new front-of-pack nutrition label in France: the five- colour Nutri-Score. Public Health Panorama. 2017. URL: https://apps.who.int/iris/bitstream/handle/10665/325207/php-3-4-712-725-eng.pdf [accessed 2022-05-30]
61. Schwarzer R, Schuz B, Ziegelmann JP, Lippke S, Luszczynska A, Scholz U. Adoption and maintenance of four health behaviors: theory-guided longitudinal studies on dental flossing, seat belt use, dietary behavior, and physical activity. Ann Behav Med 2007 Apr;33(2):156-166. [doi: 10.1007/BF02879897] [Medline: 17447868]
62. Renner B, Schwarzer R. Risk and Health Behaviors. Documentation of the Scales of the Research Project: “Risk Appraisal Consequences in Korea” (RACK). International University Bremen & Freie Universität Berlin. 2005. URL: http://www.gesundheitsrisiko.de/docs/RACKEnglish.pdf [accessed 2022-05-30]
63. Health-Care Self-Determination Theory Questionnaire Packet. Self-Determination Theory. URL: https://selfdeterminationtheory.org/health-care-self-determination-theory/ [accessed 2021-06-01]
64. King WR, He J. A meta-analysis of the technology acceptance model. Information & Management 2006 Sep;43(6):740-755. [doi: 10.1016/j.im.2006.05.003]
65. Ashton LM, Morgan PJ, Hutchesson MJ, Rollo ME, Collins CE. Feasibility and preliminary efficacy of the 'HEYMAN' healthy lifestyle program for young men: a pilot randomised controlled trial. Nutr J 2017 Jan 13;16(1):2 [FREE Full text] [doi: 10.1186/s12937-017-0227-8] [Medline: 28086890]
66. Gell NM, Wadsworth DD. The use of text messaging to promote physical activity in working women: A randomized controlled trial. J Phys Act Health 2015 Jun;12(6):756-763 [FREE Full text] [doi: 10.1123/jpah.2013-0144] [Medline: 25110303]
67. Rosner B. Fundamentals of Biostatistics. Belmont, California: Duxbury Press; 2011.
68. Compernolle S, Vandelanotte C, Cardon G, De Bourdeaudhuij I, De Cocker K. Effectiveness of a web-based, computer-tailored, pedometer-based physical activity intervention for adults: a cluster randomized controlled trial. J Med Internet Res 2015 Mar 09;17(2):e38 [FREE Full text] [doi: 10.2196/jmir.3402] [Medline: 25665498]
69. 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]
70. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009 Jun 29;338:b2393 [FREE Full text] [doi: 10.1136/bmj.b2393] [Medline: 19564179]
71. Bryman A. Social Research Methods. Oxford, UK: Oxford University Press; 2016.
72. Sequeira M, Rickenbach M, Wietlisbach V, Tullen B, Schutz Y. Physical activity assessment using a pedometer and its comparison with a questionnaire in a large population survey. Am J Epidemiol 1995 Nov 01;142(9):989-999. [doi: 10.1093/oxfordjournals.aje.a117748] [Medline: 7572981]
73. Murakami K. Nutritional quality of meals and snacks assessed by the Food Standards Agency nutrient profiling system in relation to overall diet quality, body mass index, and waist circumference in British adults. Nutr J 2017 Sep 13;16(1):57 [FREE Full text] [doi: 10.1186/s12937-017-0283-0] [Medline: 28903773]
74. Lisetti C, Yasavur U, De LC, Amini R, Rishe N, Visser U. Building an on-demand avatar-based health intervention for behavior change. In: Proceedings of the 25th International Florida Artificial Intelligence Research Society Conference, FLAIRS-25. 2012 Presented at: 25th International Florida Artificial Intelligence Research Society Conference; May 23-25, 2012; Marco Island, FL.
75. Sisson SB, Camhi SM, Church TS, Tudor-Locke C, Johnson WD, Katzmarzyk PT. Accelerometer-determined steps/day and metabolic syndrome. Am J Prev Med 2010 Jun;38(6):575-582. [doi: 10.1016/j.amepre.2010.02.015] [Medline: 20494233]
76. Oftedal S, Holliday E, Attia J, Brown W, Collins C, Ewald B, et al. Daily steps and diet quality, but not sleep, are related to mortality in older Australians. Journal of Science and Medicine in Sport 2019 Oct;22:S53-S54. [doi: 10.1016/j.jsams.2019.08.244]
77. Moroshko I, Brennan L, O'Brien P. Predictors of dropout in weight loss interventions: a systematic review of the literature. Obes Rev 2011 Nov;12(11):912-934. [doi: 10.1111/j.1467-789X.2011.00915.x] [Medline: 21815990]
78. Orfanos P, Naska A, Trichopoulou A, Grioni S, Boer JMA, van Bakel MME, et al. Eating out of home: energy, macro- and micronutrient intakes in 10 European countries. The European Prospective Investigation into Cancer and Nutrition. Eur J Clin Nutr 2009 Nov 4;63 Suppl 4(S4):S239-S262. [doi: 10.1038/ejcn.2009.84] [Medline: 19888277]

79. Teixeira PJ, Palmeira AL, Vansteenkiste M. The role of self-determination theory and motivational interviewing in behavioral nutrition, physical activity, and health: an introduction to the IJBNPA special series. Int J Behav Nutr Phys Act 2012 Mar 02;9(1):17 [FREE Full text] [doi: 10.1186/1479-5868-9-17] [Medline: 22385644]
80. Alahmari AD, Mackay AJ, Patel AR, Kowlessar BS, Singh R, Brill SE, et al. Influence of weather and atmospheric pollution on physical activity in patients with COPD. Respir Res 2015 Jun 13;16(1):71 [FREE Full text] [doi: 10.1186/s12931-015-0229-z] [Medline: 26071400]
81. Marktanteile der Top 10 führenden Detailhändler in der Schweiz im Jahr 2020. Statista. 2020. URL: https://de.statista.com/statistik/daten/studie/368538/umfrage/marktanteile-der-fuehrenden-unternehmen-im-lebensmittelhandel-in-der-schweiz/ [accessed 2021-06-01]

Abbreviations

FSA-NPS DI: British Food Standards Agency Nutrient Profiling System Dietary Index
HAPA: Health Action Process Approach
mHealth: mobile health
NCD: noncommunicable disease
RCT: randomized controlled trial
WHO: World Health Organization

Edited by R Kukafka; submitted 29.07.21; peer-reviewed by J Liang, J Ryan; comments to author 18.09.21; revised version received 03.11.21; accepted 20.04.22; published 07.07.22.

Please cite as:
Mönninghoff A, Fuchs K, Wu J, Albert J, Mayer S
The Effect of a Future-Self Avatar Mobile Health Intervention (FutureMe) on Physical Activity and Food Purchases: Randomized Controlled Trial
J Med Internet Res 2022;24(7):e32487
URL: https://www.jmir.org/2022/7/e32487
doi:10.2196/32487
PMID:35797104

Original Paper

Pilot Implementation of a User-Driven, Web-Based Application Designed to Improve Sexual Health Knowledge and Communication Among Young Zambians: Mixed Methods Study

Anjali Sharma[1], BA, SCD; Chanda Mwamba[1], MPhil; Mwila Ng'andu[1], MSc; Vikwato Kamanga[1], BSc; Mayamiko Zoonadi Mendamenda[1], BSc; Yael Azgad[2], MA; Zainab Jabbie[1], BSc; Jenala Chipungu[1], MPH; Jake M Pry[1,3], MPH, PhD

[1]Centre for Infectious Disease Research in Zambia, Lusaka, Zambia

[2]Avert, London, United Kingdom

[3]Department of Public Health Sciences, University of California, Davis, CA, United States

Corresponding Author:
Jake M Pry, MPH, PhD
Department of Public Health Sciences
University of California
Medical Sciences 1-C
One Shield's Ave
Davis, CA, 95616
United States
Phone: 1 9366616885
Fax: 1 530 752 3239
Email: Jmpry@ucdavis.edu

Abstract

Background: Digital health interventions show promise in improving the uptake of HIV services among adolescents and young people aged 15 to 24 years in sub-Saharan Africa.

Objective: This study aimed to pilot-test a theory-based, empirically grounded web-based application designed to increase condom-related knowledge, sexual and reproductive health (SRH) communication, and healthier choices among young Zambians.

Methods: We conducted a pre-post quasi-experimental evaluation of the user-driven *Be in the Know Zambia* (BITKZ) web application using web-based surveys and in-depth interviews (IDIs) on the phone. We enrolled participants using social media advertisements. Our final analysis set comprised 46.04% (749/1627) of participants in the intervention group (which received the BITKZ link) and 53.96% (878/1627) of participants in the comparison group (no intervention). We collected survey data at study enrollment (baseline) and 5 weeks after the first enrollment in each group. Approximately 85% (637/749) of BITKZ users completed a user survey, of whom 9.3% (59/637) participated in IDIs. We calculated the time interfacing with BITKZ using the application log files. We conducted descriptive analyses to describe baseline characteristics and the user experience. At the endline, we assessed association using a *t* test and adjusted logistic regression for binary outcomes and ordinal regression for ordered outcomes, conditioning on age, sex, marital status, and employment status. We used adjusted average treatment effects (aATE) to assess the effects of BITKZ intervention. We conducted rapid matrix analyses of IDI transcripts in Microsoft Excel, sorting the data by theme, gender, and experience rating.

Results: Users rated BITKZ highly (excellent: 352/609, 57.8%; good: 218/609, 35.8%). At the endline, the intervention group had a higher level of knowledge related to condoms (adjusted odds ratio [aOR]: 1.35, 95% CI 1.06-1.69) and on wearing condoms correctly (aOR: 1.23, 95% CI 1.02-1.49). Those who had full-time employment had increased odds of knowing how to wear condoms correctly (aOR: 1.67, 95% CI 1.06-2.63) compared with those who reported being unemployed, as did men when compared with women (aOR: 1.92, 95% CI 1.59-2.31). Those in the intervention group were more likely to score higher for intention to test for sexually transmitted infections (STIs; aATE 0.21; *P*=.01) and HIV (aATE 0.32; *P*=.05), as well as for resisting peer pressure (aATE 2.64; *P*=.02). IDIs corroborated increased knowledge on correct condom use among men and female condoms among women, awareness of STIs and testing, and resistance to peer pressure. Interviewees provided examples of more open SRH communication with partners and peers and of considering, adopting, and influencing others to adopt healthier behaviors.

Conclusions: Despite the high baseline awareness of SRH among Zambian adolescents and young people with internet access, BITKZ provided modest gains in condom-related knowledge, resistance to peer pressure, and intention to test for STIs and HIV.

(J Med Internet Res 2022;24(7):e37600) doi:10.2196/37600

KEYWORDS

sexual and reproductive health; web application; digital health intervention; pilot study; quasi-experiment; adolescent; young people; Zambia; sub-Saharan Africa; mobile phone

Introduction

The current rate of decline in HIV incidence among adolescents and young people aged 15 to 24 years is insufficient to end the AIDS epidemic by 2030 [1]. Southern and East African regions, although performing well, are projected to achieve an 84% reduction in 2050 from 2010 HIV incidence rates among adolescents and young people [1], which is below the 90% reduction target for 2030 [2]. Although overall HIV incidence has decreased in these regions, disproportionate numbers of urban young women aged 15 to 24 years and young men aged 20 to 29 years [3] are newly infected with HIV, albeit with some country-level variations [4]. The COVID-19 pandemic further complicated health service delivery and threatens to reverse decade-long gains in HIV prevention among adolescents and young people in sub-Saharan Africa (SSA) [5,6].

Digital health interventions (DHIs) offer an opportunity to reach digitally connected adolescents and young people with sexual and reproductive health (SRH) information, education, and services during the COVID-19 pandemic restrictions [7,8]. An age-unrestricted meta-analysis of SSA data found that DHIs improved HIV prevention knowledge and intention to act but not attitudes or perceived self-efficacy [9]. In addition, DHIs without human interaction showed no effect on the uptake of HIV prevention behaviors [9]. Interactive DHIs provide knowledge and tailored personalized feedback to support emotions, decision-making, and behavior change. Another meta-analysis, including a Zambian and Ugandan study and 12 studies among adolescents and young people, found positive effects of interactive DHIs on HIV prevention knowledge, intention, and behavior [10]. Finally, a meta-analysis of SRH DHI among adolescents and young people showed increased condom use, reduced sexual intercourse, and mixed results for improved knowledge but did not include any studies from SSA [11]. These mixed and limited results suggest the need for more research on the effect of DHIs on SRH among adolescents and young people across SSA [8,9,11].

Although a comparatively low proportion of adolescents and young people live with HIV in Zambia [12], HIV prevalence has increased among urban young men aged 15 to 24 years [13], and young women aged 17 to 19 years remain at a precipitous risk for new HIV infections [14]. Progress toward reducing HIV incidence in this population remains suboptimal [15]. The COVID-19 pandemic has further disrupted government-endorsed, school-based, and interpersonal structural, behavioral, and combination HIV prevention efforts and may have increased reliance on mass and social media platforms [15-18]. We sought to add to these efforts by designing a user-driven DHI incorporating informational elements based on the expressed needs and preferences of adolescents and young people during the containment periods of the first 2 COVID-19 waves in Zambia [19]. In this paper, we present the effect of the pilot implementation of the cocreated *Be in the Know Zambia* (BITKZ) web-based application on SRH knowledge, communication, intentions, and behavior. This study adds to the growing body of literature on the measurement of the effect of DHI on SRH among adolescents and young people in SSA [9,20-23].

Methods

We conducted a pre-post quasi-experimental mixed methods evaluation using web-based survey data collection on Qualtrics (Qualtrics International Inc) [24] and in-depth interviews (IDIs) on the phone from June to August 2021.

Enrollment

To target persons eligible for the pilot study, we placed advertisements on Facebook (Meta Platforms Inc) and contacted established Zambian adolescent groups in youth-friendly SRH spaces identified in formative phase interactions and followed up using WhatsApp (Meta Platforms Inc) groups or email. WhatsApp is a commonly used mobile-based SMS text message app. Individuals aged 18 to 24 years, living in Zambia, and able to understand and give informed consent in English (the language of the app) were eligible. Those who did not meet all the eligibility criteria or declined to consent were excluded.

Eligible individuals were required to provide consent for data storage and analyses for publication purposes and to be contacted by phone or email. In addition, those interested in participating in the study were required to provide a phone number or email address to receive a link to a secure baseline and endline survey and to receive airtime on completing each survey.

The first 1500 individuals were targeted for the intervention group and were consecutively enrolled if they were eligible and agreed to join the BITKZ application for at least 1 month and complete the user and endline surveys. The next 1500 individuals were consecutively enrolled into the comparison group if they were eligible and agreed to participate in the endline survey after a month of enrollment. We decided to enroll adolescents and young people sequentially rather than in parallel to ensure an adequate sample size in the intervention arm, given that there was no precedent for this type of web-based enrollment, intervention delivery, and evaluation in Zambia.

Those enrolled in the intervention group were also offered a chance to be invited to an IDI on either (1) their user experience and interaction with the application or (2) their user experience

with each feature of the BITKZ app. Individuals who agreed to be contacted for an IDI were asked for their preferred contact information and consent to audio recording and transcription.

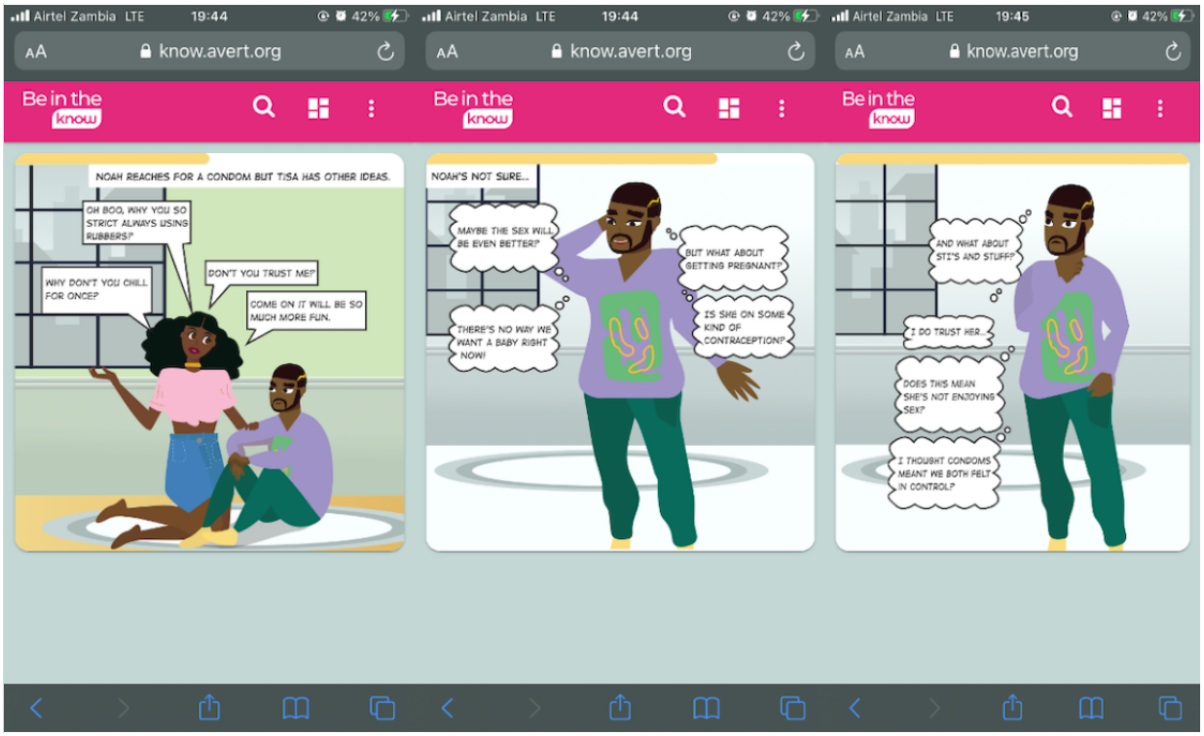
Intervention

Avert, a UK-based organization, engineered the BITKZ internet or web-based application to provide SRH content for young Zambians aged 18 to 24 years. Avert uses innovative digital health approaches to help adolescents and young people make healthier SRH choices to reduce adverse SRH outcomes and ultimately improve their lives. The application content and design features were based on insights from the formative phase from August 2020 to May 2021, including qualitative phone-based interviews with 18 Zambian adolescents and young people from August to October 2020 and cocreation sessions held with 51 Zambian adolescents and young people from September to December 2020 on Facebook (3 groups of 7 men, 9 women, and 6 both) and WhatsApp (3 groups of 9 men, 7 women, 13 both). These groups further helped refine the BITKZ functionality through user feedback collected between March and May 2021.

BITKZ is grounded in the practical application of theory-based methods [25] and builds on the foundational work collected during the initial formative phase. During the formative work, we discovered that adolescents and young people did not know how to use condoms, had unplanned sex under the influence of peers, and longed to receive guidance from knowledgeable and trusted sources. Thus, BITKZ provided SRH information to adolescents and young people to increase condom-related knowledge, resistance to peer pressure, and SRH communication with people who matter to support the outcomes of interest to Avert—intention and practice to prevent sexually transmitted infections (STIs), HIV, and unintended pregnancies. Multimedia Appendix 1 demonstrates how BITKZ used the taxonomy provided by Kok et al [25] to map the intervention to the behavior we intended to change using levers that responded to the formative work while being pragmatic because of the limited time, resources, and communication platforms available for the project. BITKZ tailored SRH information to include imagery, language, and context relevant to the adolescents and young people in Zambia, setting up scenario-based risk information to encourage dialog, active learning, and social support. These scenarios, along with positive and gain-based messages, aimed to promote implementation intentions and goal setting.

In application, BITKZ appealed to personal identity [26] by inviting players to choose among 3 male and 3 female characters aged 18 to 24 years, with distinctive traits (modern independent go-getters, sophisticated dreamers, or religious loners) developed from narratives gathered during the formative phase of development. Users would then enter a comic strip where each male and female character faces dilemmas centered on 3 themes—condoms, preventing pregnancy, and staying healthy—aiming to provoke dialog and reflection [27]. Users could interact with these 3 themes through 6 features, the screenshots of which are depicted in Multimedia Appendix 2 to illustrate how they might aid adolescent and young people users in SRH-related communications. Briefly, the 6 features included the comic strips modeling dilemmas using real-life scenarios described previously (Figure 1), visual guides on how to use various condoms and contraceptives, sharable *Let's Talk* and *Top Tips* cards to support users to think through choices and provide action-focused ideas, frequently asked questions about SRH, and quizzes.

Figure 1. Screenshot of the scenario.



These features aimed to positively reinforce knowledge and conversations exploring choices on topics that concern the user [28]. In addition, the application had a gamification feature [29] where users could win up to 5 web-based badges for

commenting, sharing their favorite *Top Tips* and *Let's Talk cards* using BITKZ's integrated share function, looking for answers to frequently asked questions, reading all the visual guides, and completing all quizzes [30]. Given the variety of media with which participants might interact, we sought here only to assess any uptake compared with no uptake of or access to BITKZ materials.

Following the baseline survey on Qualtrics [24], participants in the intervention group could follow the application content in any order and frequency on any electronic device with an internet connection during the 3-month intervention period.

Data Collection

Web-based Pre-Post Intervention Surveys

All surveys were developed and administered on Qualtrics [24], and reimbursements (airtime) were made through cGrate (Zambia Ltd) [31]. All participants enrolled and self-administered the baseline web-based survey from June 7 to August 3, 2021. On completion, they received 30 Zambian Kwacha (ZMW; US $1.50). Intervention participants who completed the baseline survey received a single-user link to the BITKZ application on email or WhatsApp (Meta Platforms Inc) as per their preference. Participants in the comparison group were advised to seek SRH advice as usual. Participants in the intervention group received a link to a self-administered web-based user survey a week after accessing the application. Those who completed the user surveys received 20 ZMW (US $1.00).

Intervention and comparison participants received an invitation 5 weeks after enrollment to complete the endline survey followed by biweekly reminders by WhatsApp (Meta Platforms Inc) or email. All participants self-administered the endline survey on the web from July 12 to September 2, 2021. On completion, they received 50 ZMW (US $2.50).

Phone IDIs

From July to August 2021, we conducted phone interviews with 59 participants who completed the endline and user surveys (Table 1).

We sequentially telephoned participants from among those who rated the BITKZ application as excellent, good, and poor, ensuring a gender balance. We interviewed 20% (12/59) of individuals each among those who rated the BITKZ application as *good* and *excellent* (replacing 2 individuals who missed 2 consecutive appointments). Of the 9 individuals who rated the application as *poor*, 8 (89%) agreed to the interview. We sequentially telephoned additional intervention participants to collect feedback on each application feature. Of the 78 participants who were telephoned, we interviewed a total of 27 (35%; n=13, 48% women, and n=14, 52% men) intervention participants on application features that stood out prominently for them to obtain at least six views on each feature. The remaining 86% (51/59) of participants did not answer, could not recall a feature, or recalled a feature already discussed by 6 interviewees.

We confirmed the identity of the person on the phone by name and age, after which we further asked whether they remembered being enrolled and interacting with the BITKZ application. If not, we thanked them for their time and went on to the next person on our list. We reminded others about their web-based agreement and confirmed their consent to the IDI, audio recording, and transcription. All interviews explored the effects of the application on their sexual health knowledge, communication, and behavior. Interviewees received 100 ZMW (US $5.00) reimbursement by cGrate [31].

Table 1. Postintervention in-depth interviews on BITKZ[a] perception and experience (N=59).

Criteria	Men, n (%)	Women, n (%)	Total, N
BITKZ rating			
Excellent	6 (50)	6 (50)	12
Good	6 (50)	6 (50)	12
Poor	4 (50)	4 (50)	8
Prominent features	14 (52)	13 (48)	27

[a]BITKZ: Be in the Know Zambia.

Survey Instruments and Measures

All measures were collected from the intervention and comparison groups using the same survey instrument at baseline and endline, which can be found in Multimedia Appendix 3.

The primary outcomes for all participants included condom-related knowledge, ability to talk to people who matter, and frequency of seeking SRH advice from them in the past month. Condom knowledge included 6 items (true or false) and the identification of 6 correct steps when using a condom from a list of 14 options adapted from Stanton et al [32]. Correct responses were coded as 1, and all other responses were coded as 0 to derive the mean and percentage of correct responses.

The ability to seek advice on SRH included a list of 10 potential options for people who matter in the respondent's life and was measured on a Likert scale ranging from definitely, probably, probably cannot, and definitely cannot (Cronbach α=.80) [33]. The frequency of seeking SRH advice by source in the past for month more than once a week, once a week, 1-2 times a month, and never) was calculated per source (Cronbach α=.86).

Other secondary outcomes measured the intention to use condoms at the next sexual intercourse (not at all, somewhat, and very likely). For a subpopulation of sexually active participants, we also asked about the intention to test for STIs if symptomatic and for HIV in the next 6 months. Higher scores reflected higher occurrence or likelihood of events.

We also collected participant characteristics, including self-reported age, gender, marital status, education level, perceived socioeconomic status relative to their community (rich-poor and very-not respected), and employment status. We asked about sexual debut (never had sex; true or false), currently sexually active (past 6 months; true or false), condom use at the last sexual intercourse (true or false), unplanned last sexual intercourse (true or false) [34], and most recent tests for STIs and HIV (<3 months, 3-6 months, 6-12 months, >12 months, and never). The condom knowledge questions correspond to questions 3.1 to 3.6 (Multimedia Appendix 3).

Other independent variables were drawn from sociocognitive theories [27] to include scales to measure permissive attitude (Cronbach α =.68), sexual norms (Cronbach α=.60), and self-efficacy (Cronbach α=.80), derived from a study by Muhammed et al [35] (Multimedia Appendix 3, question 4); frequency of being pressured into making unhealthy choices (every time, sometimes, rarely, and never) derived from the Kaiser National survey (Multimedia Appendix 3, question 5; Cronbach α=.80) [36]; and confidence to get a condom and STI or HIV test whenever wanted (very, somewhat, and not at all confident) [37]. Although we did not expect to see changes in these variables, we considered them as possible influencers of variables measuring intention and practice.

Application Experience and Use

We gathered ratings of overall participant application experience (excellent, good, and poor), as well as perceived aesthetics; engagement; functionality; reaction to the information; and the likelihood of sharing, continuing to use, and willingness to pay from the user survey (very, quite, and not at all).

The app software automatically generates a user log file that records each user interaction, which is time-stamped. These data were downloaded from the application's website using Amplitude and the Amplitude export application programming interface.

Topics Covered by IDIs

Interviews with the first 32 participants explored the overall user experience, how the application content was used or had influenced the user, and recommendations for improvement. The remaining 27 IDIs explored experience with each of the six features: ease, aesthetics, entertainment, learning, sharing, change in SRH practices, and recommendations for improvement.

Data Analysis

Quantitative Analysis

We conducted descriptive analyses to determine intervention status comparability and assess differences in participants' sociodemographic characteristics, SRH norms, SRH behaviors, and communication about relationships and SRH, as well as to describe user experience.

We calculated a regression-adjusted average treatment effect (aATE) on SRH knowledge, permissive attitude, sexual norms, self-efficacy, peer pressure, and confidence in procuring STI tests and condoms, as well as intention to test for STI and HIV and use condoms at the next intercourse. Scores for permissive attitudes and sexual norms were based on the sum of positive SRH-competent responses to question sets (Multimedia Appendix 3, question 4). Statistical significance of differences was assessed using a *t* test, comparing mean differences between comparison and intervention at the endline, where appropriate. Regression analysis was used to assess the impact of BITKZ on condom-related knowledge. All multivariate impact analyses incorporated an intention-to-treat approach and controlled for sociodemographic characteristics (age, sex, marital status, educational attainment, employment status, perceived wealth, and perceived respectability at baseline), baseline measures of outcomes, and intervention status. All primary outcome analyses were completed using Stata (version 16.1) [38], and the duration per user per day was calculated using R software and the *dplyr* package [39].

Application Experience and Use

We analyzed participants' use to assess the median time spent on the application. We gathered and analyzed the participants' application experience (excellent, good, and poor) from the user survey.

Phone IDIs

We conducted a rapid matrix analysis [40] of IDIs to extract the overall user experience and self-reports with examples of using the application to communicate; its influence on knowledge, resistance to peer pressure, intention to communicate with people who matter, and use of condoms and test for STI and HIV; and condom use and STI and HIV testing. Outcomes were iteratively compared across transcripts, categorized, and synthesized into analytical summaries for interpretation.

Ethics Approval

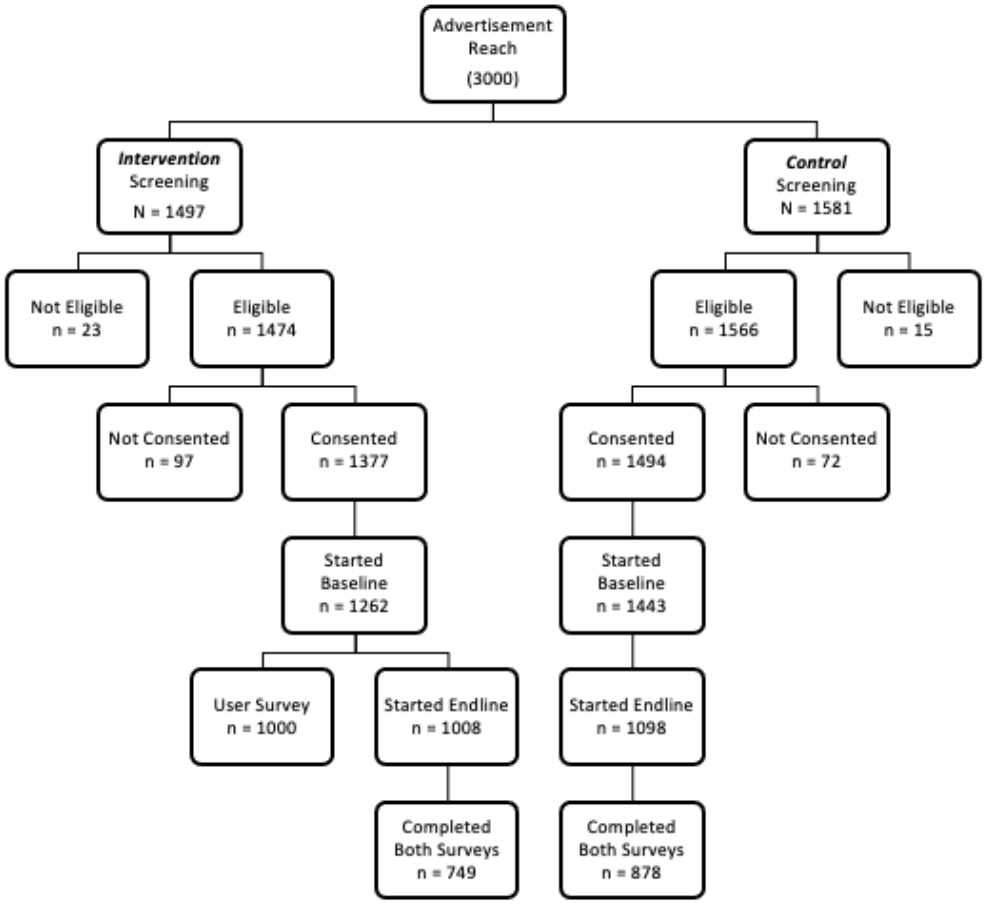
The University of Zambia Biomedical Research Ethics Committee approved the study (institutional review board protocol number 811-2020), and the National Health Research Authority granted the authority to conduct the research.

Results

Enrollment

The study targeted 3000 participants for the pilot study (Figure 2). Individuals who clicked on the recruitment advertisement were sent to a Qualtrics [30] webpage with a 4-question eligibility screener. We registered 3078 screener attempts, 98.77% (3040/3078) eligible records, 94.44% (2871/3040) consents, 94.44% (2705/2871) baselines of 16 questions (Multimedia Appendix 3), and 73.35% (2106/2871) endline surveys. The final sample comprised 1627 participants who completed both the baseline and endline surveys. We removed 230/1627 (14.1%) duplicates and ineligible accounts from the final data set. We observed a low attrition rate from screening to enrollment (2871/3078, 6.7%) and a reasonable completion rate (1627/2871, 56.7%).

Figure 2. Participant flow.



Participant Characteristics

Most of the 1627 individuals were single (1502/1613, 93.12%), with a median age of 22 (IQR 21-23) years and from averagely wealthy (868/1340, 64.78%) and respected (774/1293, 59.86%) families. Approximately half were men (822/1613, 50.96%), college educated (811/1615, 50.22%), and trainees or students (706/1614, 43.74%).

The comparison group had a higher proportion of college or university-trained individuals (462/872, 52.9%, vs 349/743, 46.9%; *P*=.002) and in the student or trainee category (424/872, 48.6%, vs 282/742, 38%; *P*<.001) than the intervention group (Multimedia Appendix 4).

The 2 groups were similar with regard to sexual health history, with 42.33% (682/1611) testing for HIV and 63.61% (1014/1594) testing for STIs (Multimedia Appendix 4, Table S3). The intervention group had a higher proportion of sexually active individuals (404/564, 76.0%, vs 396/521, 71.6%), who had sex in the past 6 months (294/510, 57.6%, vs 293/565, 51.9%), with slightly lower condom use (281/369, 76.2%, vs 317/402, 78.9%) and slightly more unplanned sex (194/369, 52.6%, vs 221/403, 54.8%) at the last sexual intercourse.

Adjusted Average Treatment Effect for BITKZ Intervention

Overview

The intervention group was more likely to score higher on the intention to test for STI (aATE: 0.21; *P*=.01) and HIV (aATE: 0.32; *P*=.05) than the comparison group (Multimedia Appendix 4). The effect measures for resisting peer pressure increased more than 2-fold (*P*=.02), likely because of heterogeneity in responses and higher possible scores. No other statistically significant effects were observed.

The intervention group had a modest increase in the likelihood to score higher in SRH knowledge (aATE: 0.12), self-efficacy (aATE: 0.21), use of condoms at the next sexual intercourse (aATE: 0.19%), and confidence in getting an STI test (aATE: 0.18) and condoms (aATE: 0.29) when needed. The likelihood of scoring higher for permissive attitudes was 91% higher in the intervention group, although they were 8% less likely to score higher for sexual norms.

Intention to Test for STI and HIV and Use Condoms

Those aged 20 years had significantly higher adjusted odds of testing for STIs if symptomatic (adjusted odds ratio [aOR] 2.5, 95% CI 1.3-4.81) than those aged 18 years. Higher odds of HIV testing were observed among those aged 19 years (aOR 1.73, 95% CI 1.02-2.92), 20 years (aOR 1.7, 95% CI 1.01-2.86), and 23 years (aOR 1.81, 95% CI 1.14-2.91) years than among those aged 18 years. Those reporting average perceived wealth had

lower odds of testing for STIs if symptomatic (aOR 0.69, 95% CI 0.52-0.93) and testing for HIV (aOR 0.68, 95% CI 0.54-0.86) than those identifying with above-average perceived wealth.

In the IDIs, more women than men expressed intent to test for STI and HIV, motivated by self-preservation and to secure their future. For example, a woman aged 22 years claimed the following:

> *To engage in sex, it has to be a safe one, so that I can continue with my education and my career.*

Both men and women provided examples of taking up STI or HIV testing as they felt more knowledgeable, motivated, and self-confident, as encapsulated in the following quote:

> *We are, you know, ignorant about sex...before I used the application, I used to be frustrated that "Ok, I slept with that girl but ah!" I start asking questions "Maybe I get STIs and whatnot." But once I started using the App I got motivated, I went for HIV testing, STIs like that, so that I am updated. Then issues of how to use a condom used to be a challenge. I didn't know how, if I am about to have sex, how I am supposed to put on the condom or remove it [I: hmm] So when I started using the application, I know a lot. Yeah I know a lot. I became motivated and I even went for HIV testing.* [Male, 22 years]

Many participants felt nudged by the knowledge that "not all STIs have symptoms" and "that STIs can be treated. So, if at any time I find out that I have STIs, I don't have to be worried...I can get help from a health clinic." They thought that having such information helped those who "may say they feel stigmatized when they have that. But if they use this app, they will know their rights, and they'll be able to learn and also to get tested for STIs whenever they have symptoms."

Peer Pressure

The intervention group had a slightly higher mean score (19.81, 95% CI 19.56-20.06) than the comparison group (19.68, 95% CI 19.44-19.91) for the ability to resist peer pressure. In the IDIs, both men and women reported gaining insight into how peer pressure can lead to undesirable outcomes and reported having an internal debate to decide their course of action and learning strategies to resist peer pressure. Multimedia Appendix 5 contains illustrative quotes to show the effect of BITKZ on the ability to resist pressure, condom and contraceptive-related knowledge, and partner communication in BITKZ users' own words.

Condom Knowledge

At the endline, the level of condom-related knowledge was 35% higher (aOR 1.35, 95% CI 1.07-1.69) among those who received the intervention than among those in the control group (Multimedia Appendix 4). In addition, men had a 27% lower level of condom-related knowledge than women (aOR 0.73, 95% CI 0.58-0.92).

Conditional or matched logistic regression comparing baseline to endline by intervention status indicated significantly increased odds of improvement in knowledge regarding the need to try different types of condoms to suit both partners in both groups, more so among intervention (matched odds ratio 3.35, 95% CI 2.5-5.33) than comparison participants (matched odds ratio 2.03, 95% CI 1.34-3.08).

Ordinal regression on correct condom use indicated increased odds of scoring higher (0-6 possible) in knowledge on how to wear the condom among the intervention group (aOR 1.27, 95% CI 1.06-1.54) compared with the control group, among men (aOR 1.92, 95% CI 1.59-2.31) compared with women, and among those employed full time (aOR 1.67, 95% CI 1.06-2.63) compared with those reporting as unemployed (Multimedia Appendix 4).

In the interviews, young people expressed their happiness with the information about the different options for condoms that can be used during sexual intercourse available on the BITKZ application, as well as evidence of learning about contraceptive choices (Multimedia Appendix 5). All participants emphasized that an important lesson they had picked from the application visual guides was the correct way of using condoms:

> *We don't have someone to tell us how to use condoms to take care of ourselves we are just doing things blindly.*

Communication With People Who Matter About SRH Topics

In descending order, participants felt they could most definitely seek advice from people who matter, defined as health care workers, boyfriends or girlfriends, friends, peers, other adults, teachers, siblings, parents, community leaders, and priests (Multimedia Appendix 4). Comparison group participants were more definite about their ability to seek advice or ideas on SRH from a priest (mean 2.4, SD 1.16 vs mean 2.52, SD 1.15; *P*=.04).

At baseline, 37.9% (318/838) of comparison group participants reported seeking advice from the church on SRH in the past month compared with 31.4% (227/724) among those in the intervention group (*P*=.001), a difference that persisted into the endline (260/741, 49.1%, vs 355/848, 35.1%; *P*=.05). At the endline, more comparison group than intervention group participants reported seeking SRH advice in the past month from family (502/844, 59.5%, vs 387/740, 52.3%; *P*=.02) and a professor or teacher (435/846, 51.4%, vs 335/742, 45.1%; *P*=.03).

Intervention group participants regarded talking to adults or parents about SRH as taboo because of moral censure, religious beliefs, sociocultural norms, and attitudes regarding adolescent sexual behavior. According to them, adults thought that "there are things adolescents should not know at that age"; otherwise "they may want to practice sex." In addition, participants indicated that parents did not discuss condom use and cautioned them to "take it easy in life, there are a lot of diseases and do not misbehave."

Young women in the intervention group disclosed that they felt generally *uncomfortable* and *shy* to have open conversations about sex. They felt less confident in their ability to discuss and insist on condom use with their partner, leaving them worried after having sex. Some participants explained that the application boosted their confidence in discussing safe sex with their

partners. The relatable scenarios and discussion starters on the application provided guidance on how they could raise topics about condom use with sexual partners. Some women considered it important to have the application so that they could invite their partners to use it and open up discussions about safe sex (Multimedia Appendix 5).

Both young men and women reported that their friends commonly provided information that was often not complete or correct. All participants noted that the application created an environment for *smart knowledge* and *confidence* when discussing sex and SRH with friends. Many participants described recent efforts to share information gathered from the application with their friends and extended invitations to friends to access the application and act on its recommendations:

> *Yeah like I shared about the App, we read some stories and answered questions with about eight girls at my boarding house. So we later went to get condoms from the clinic, from our nearest clinic.* [Female, 21 years]

> *I had a conversation with one of my friends about using condoms [...], he said, "Now I can't take action right now because those things [condoms] are too expensive!" That's when I told him that, "My friend, these things, it's not all about buying. You can just go to the clinic, you take." That's the action which he took. He went to the clinic and he was given those things.* [Male, 23 years]

User Experience

The application captured linking data for 80.2% (601/749) of intervention participants, half of whom spent 17 (IQR 6-48) minutes on BITKZ. Among the intervention group, 85% (637/749) completed the user survey and rated BITKZ highly (excellent: 359/637, 56.4%; good: 237/637, 37.2% good). Most items under the constructs of aesthetic, engaging, functional, useful, and shareable were rated *very* by >60% of the respondents (Multimedia Appendix 4). Exceptions included the ability to personalize content (*very*; 302/637, 47.4%), likelihood of sharing if in own language (*very*; 360/637, 56.5%), vastness of content (*very*; 328/637, 51.5%), and willingness to pay (*very*; 205/637, 32.2%), although the proportion rating these items as *not at all* was <11%. IDI participants who rated the app *poor* explained that they had problems with functionality because of bandwidth issues.

Discussion

Principal Findings

The BITKZ application met its objective of increasing condom-related knowledge and resistance to peer pressure among young Zambians aged 18 to 24 years who use the internet. The BITKZ application did not achieve a statistically significant increase in communication about SRH, although BITKZ users described using it to inform and motivate people who matter to adopt healthier sexual choices. The intervention group had a higher intent to test for STIs and HIV, possibly as the application reduced stigma and fear associated with prolonged morbidity and infectiousness [41-44]. Although not statistically significant, possible modest improvements in sexual norms, self-efficacy, confidence to procure a condom, STI test, and intent to use condoms in the next sexual intercourse suggest that the BITKZ application may influence young adults to plan and practice safer sex [45,46]. Triggers and prompts that encourage site visits, use, and engagement among adolescents and young people need further investigation [47].

Our evaluation confirmed the need to include visual guides on correct condom use in comprehensive sexuality education offered in Zambia [48,49]. However, more research is needed on the effectiveness of educational materials using drawings of disembodied parts, particularly for communicating the correct use of female condoms [50]. The Zambian legal framework for the depiction of full images of men and women for sexual education needs clarification, given the prohibition of possessing *obscene* materials [51] and of transmitting them electronically [52]. This will require engagement and negotiation with the community and influential leaders; research ethics boards; regulatory bodies; the Zambia Information, Communication, and Technology Authority; and the Ministry of Justice. These and other materials can be further culturally adapted and produced remotely at low cost and at scale, as noted by the Kenyan Tumaini project, which developed and piloted a smartphone game intervention to improve SRH among young people [23].

Increased resistance to peer pressure may have been easier in our older study population than in preteens and early teenagers, as targeted by other studies [23,47]. The Tumaini project [23] addressed but did not measure the effect of peer pressure on those aged 11 to 14 years in Kenya. A digital storytelling intervention conducted in South Africa [47] found that exposure to multiple alternate views led high school girls to better understand their own emotions and behavior, including with regard to peer pressure. In our study, scenarios resonated with BITKZ users, leading to introspection and the intention to resist peer pressure. BITKZ use data, including user interactions, will be further examined for evidence of unintended peer pressure on the BITKZ platform [53].

Closson et al [54] and other comprehensive reviews [55] illustrate the correlation of personal experience and norms with those of peers, as well as the differential effects of sexual norms and behavior change interventions by gender in SSA. Unlike the previous review, our intervention reflected normative changes, possibly because of resistance to peer pressure and improved communication with peers and sexual partners [54]. However, young men in our intervention group learned more about correct condom use than did women, suggesting gendered interests and loci of control [54]. We will further analyze our data, disaggregated by sex and age groups, to understand the differential effect of this intervention by drawing on the application log files to capture use and estimate dose-response.

In addition, higher sexual permissiveness but lower sexual norm scores may reflect a healthier, sex-positive attitude, which was found to be safer for STI prevention in the United States [43]. Despite these encouraging results on condom use knowledge and intention to test for STI and HIV, gender relations and lack of youth-centered approaches at clinics may bar the uptake of

STI and HIV prevention and care-seeking behaviors. These social and structural barriers have impeded STI prevention in SSA [56-60]. Digital health innovations for training SRH care providers can link adolescents and young people to needed services and allow for the social monitoring of SRH services for adolescents and young people [61,62].

Having accurate knowledge of STIs can reduce stigma and increase intention to test for STIs [41-44] and sexual confidence [63]. Although our study demonstrated increased intention, the intervention was insufficient to open conversations between adults and adolescents and young people, which may make it less likely that adolescents and young people will access STI testing [63]. Modernization in the timing and content of *sexual teachings* [64,65] and health programs that increase adolescents' and young peoples' assertiveness can increase SRH communication with parents and other significant adults, with anticipated benefits [63]. In addition, parent-based interventions show a significant association with improved condom use and parent-child sexual communication, especially when focused on young adolescents and targeted at both parents and adolescents [66]. Evidence-based choice and design of a parent-child intervention and cocreation processes would require intergenerational engagement and mediation to diminish parents' discomfort and fears about communicating with their children on sexual matters [8,23,67-69]. Intergenerational games [23,70] and digital storytelling can remove some of the barriers to parent-child communication.

Limitations

Owing to the COVID-19 pandemic restrictions, we conducted web-based recruitment, data collection, and BITKZ implementation. As a result, our sample may be more educated, technology savvy, urban, male, and wealthy than the general adolescents and young people population [71]. In addition, the COVID-19 pandemic determined our exclusively web-based BITKZ design. A more inclusive combined web-based and in-person design may deepen engagement and help identify misconceptions and adolescents and young people–specific language [9,72]. Although use data indicate good engagement, an analysis of proximity to the time of the survey, time on the application, and time on each feature is needed. Although aligned with our intention-to-treat analysis, including exposure status can help estimate the efficacy of our intervention and its relevance based on demographic characteristics.

Web-based data collection did not allow us to confirm age and residence, although all provided Zambian phone numbers to receive communication and reimbursement in ZMW and were required to complete all the screening questions so that they would not learn why they were considered ineligible. Being completely on the web, we limited the number of survey questions and could not validate the truncated scales for the theory of planned behavior constructs [34]. We did not randomize enrollees to the intervention or comparison groups, although both groups did not differ in outcome variables at baseline. Endline improvements in the comparison group suggested the influence of maturation and testing effect [73]. This, together with the high baseline scores for both groups, may have limited the power to detect small but significant changes because of BITKZ.

We cannot rule out courtesy, recall, and social desirability bias among participants responding to IDIs [73]. Owing to time, financial, and COVID-19 restrictions, we did not measure self-reported behavior changes validated by biomarkers. A longer follow-up time is required to estimate continued engagement with the application and its medium- to long-term impact on attitude, norms, intention, and behavior. In addition, comparison group enrollment was conducted in series rather than in parallel with the intervention group; however, we anticipate selection bias to be minimal and biased toward the null, given a brief enrollment period and order of enrollment (intervention and then comparison group).

Comparison With Prior Work

We found no comparable study on similar DHIs for heterosexual young people either because of differences in targeted age (<18 years) or setting (school based) in SSA [23,74,75]. Similar programs targeting young people aged 18 to 24 years come from the clinic and web-based settings in the United States [76] and university settings in the Netherlands [77]. In addition, the Dutch *Justify Your Love* study targeted heterosexual couples in a new relationship [77]. All 5 DHIs discussed herein used random assignment of participants as control and 1 to 2 intervention conditions to provide information, motivation, and problem-solving and behavioral skills in English. In addition, BITKZ, Justify Your Love [77], and the Tumaini pilot in Kenya [23] addressed communication skills.

The DHIs differed in form and content, informed by context, formative research, behavior change theories, educational tools, and cocreation or wide consultation. Similar to BITKZ, the 2 trials were persona based [23,75] and had reward features [23,75,77]. Unlike BITKZ, 3 were set in virtual worlds [23,75,77]: one designed for offline users [75] and another for Android smartphone [23] users. The 3 DHIs in SSA stipulated duration, venues, and linear flow, some with material or teaching support [23,74,75]. The CyberSenga trial in Uganda was unique in its use of adult traditional figures as role models and completion certifications [74]. Non-SSA participants received up to €20 (US $20.86) for full participation in cash [77] or gift certificates [76], whereas BITKZ participants received up to €4 (US $4.17) in incentives, which may explain the relatively modest attrition.

Differences in purpose and design choices led to differences in the reported measures. Games and gamification have been proven to improve knowledge, motivation, and engagement in learning about SRH [74], goal setting and risk avoidance [23,74], and SRH knowledge [23,74]. Unlike the non-SSA trials YouthNet [76] and Justify Your Love [77], SSA-based DHIs reported a statistically significant increase in SRH self-efficacy [23] and, although statistically nonsignificant, increased the likelihood to remain abstinent [74]. CyberSenga [78] and Justify Your Love [77] reported increased condom use unlike YouthNet [76]. The Justify Your Love intervention did not change attitude, normative beliefs, skills toward maintenance of condom use or STI testing, and the uptake of STI testing [77]. YouthNet

reported no differences in awareness of HIV or sexually transmitted disease risk and attitudes toward condom use [76].

Conclusions

This evaluation study successfully used social media to recruit adolescents and young people aged 18 to 24 years to participate in an exclusively web-based SRH program. Young Zambians with internet access have a high awareness of SRH issues. BITKZ provided modest gains in intention to test for STIs, possibly because of the novelty of this concept vis-à-vis HIV, and in correct condom use because of insufficient prior knowledge.

Acknowledgments

The Vitol Foundation, the Risk Pool Fund, and Avert supported the development and evaluation of the BITKZ application. The content herein is solely the responsibility of the authors and does not necessarily represent the official views of the Vitol Foundation, Avert, and the Center for Infectious Disease Control in Zambia. The Vitol Foundation had no role in the review and approval of the manuscript for publication. The authors wish to thank the adolescents and young adults who cocreated the app and helped with the evaluation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Be in the know Zambia objectives and theoretical basis.
[DOCX File , 18 KB - jmir_v24i7e37600_app1.docx]

Multimedia Appendix 2
Be in the know Zambia application features.
[PPTX File , 4298 KB - jmir_v24i7e37600_app2.pptx]

Multimedia Appendix 3
Be in the know Zambia survey instrument.
[DOCX File , 36 KB - jmir_v24i7e37600_app3.docx]

Multimedia Appendix 4
Quantitative tables.
[DOCX File , 47 KB - jmir_v24i7e37600_app4.docx]

Multimedia Appendix 5
Supporting quotes.
[DOCX File , 17 KB - jmir_v24i7e37600_app5.docx]

References

1. Khalifa A, Stover J, Mahy M, Idele P, Porth T, Lwamba C. Demographic change and HIV epidemic projections to 2050 for adolescents and young people aged 15-24. Glob Health Action 2019;12(1):1662685 [FREE Full text] [doi: 10.1080/16549716.2019.1662685] [Medline: 31510887]
2. UN Joint Programme on HIV/AIDS. Fast-track: commitments to end AIDS by 2030. UNAIDS. 2014. URL: http://www.unaids.org/sites/default/files/media_asset/fast-track-commitments_en.pdf [accessed 2021-09-18]
3. Risher KA, Cori A, Reniers G, Marston M, Calvert C, Crampin A, ALPHA Network. Age patterns of HIV incidence in eastern and southern Africa: a modelling analysis of observational population-based cohort studies. Lancet HIV 2021 Jul;8(7):e429-e439 [FREE Full text] [doi: 10.1016/S2352-3018(21)00069-2] [Medline: 34197773]
4. Birdthistle I, Tanton C, Tomita A, de Graaf K, Schaffnit S, Tanser F, et al. Recent levels and trends in HIV incidence rates among adolescent girls and young women in ten high-prevalence African countries: a systematic review and meta-analysis. Lancet Glob Health 2019 Nov;7(11):e1521-e1540 [FREE Full text] [doi: 10.1016/S2214-109X(19)30410-3] [Medline: 31607465]
5. Roberts L. How COVID is derailing the fight against HIV, TB and malaria. Nature 2021 Sep;597(7876):314. [doi: 10.1038/d41586-021-02469-8] [Medline: 34508245]
6. Brown LB, Spinelli MA, Gandhi M. The interplay between HIV and COVID-19: summary of the data and responses to date. Curr Opin HIV AIDS 2021 Jan;16(1):63-73 [FREE Full text] [doi: 10.1097/COH.0000000000000659] [Medline: 33186229]

7. Wadham E, Green C, Debattista J, Somerset S, Sav A. New digital media interventions for sexual health promotion among young people: a systematic review. Sex Health 2019 Apr;16(2):101-123. [doi: 10.1071/SH18127] [Medline: 30819326]
8. Daher J, Vijh R, Linthwaite B, Dave S, Kim J, Dheda K, et al. Do digital innovations for HIV and sexually transmitted infections work? Results from a systematic review (1996-2017). BMJ Open 2017 Nov 03;7(11):e017604 [FREE Full text] [doi: 10.1136/bmjopen-2017-017604] [Medline: 29101138]
9. Manby L, Aicken C, Delgrange M, Bailey JV. Effectiveness of eHealth interventions for HIV prevention and management in Sub-Saharan Africa: systematic review and meta-analyses. AIDS Behav 2022 Feb;26(2):457-469 [FREE Full text] [doi: 10.1007/s10461-021-03402-w] [Medline: 34427813]
10. Bailey JV, Wayal S, Aicken CR, Webster R, Mercer CH, Nazareth I, et al. Interactive digital interventions for prevention of sexually transmitted HIV. AIDS 2021 Mar 15;35(4):643-653 [FREE Full text] [doi: 10.1097/QAD.0000000000002780] [Medline: 33259345]
11. Tarnate PS, Serrano KR, Torres-Ticzon VM, Senen KA. Effectiveness of digital media technologybased interventions on HIV & STI risk reduction among young people: a metaanalysis. Pediatr Infect Dis Soc Philippines J 2019;20(1):3-14 [FREE Full text]
12. Towards a global HIV prevention coalition and road map. UNAIDS. 2017. URL: https://www.unaids.org/sites/default/files/media_asset/towards-global-HIV-prevention-coalition-and-road-map_en.pdf [accessed 2021-09-26]
13. Nakazwe C, Michelo C, Sandøy IF, Fylkesnes K. Contrasting HIV prevalence trends among young women and men in Zambia in the past 12 years: data from demographic and health surveys 2002-2014. BMC Infect Dis 2019 May 17;19(1):432 [FREE Full text] [doi: 10.1186/s12879-019-4059-3] [Medline: 31101081]
14. Shanaube K, Macleod D, Chaila MJ, Mackworth-Young C, Hoddinott G, Schaap A, et al. HIV care cascade among adolescents in a "test and treat" community-based intervention: HPTN 071 (PopART) for youth study. J Adolesc Health 2021 Apr;68(4):719-727 [FREE Full text] [doi: 10.1016/j.jadohealth.2020.07.029] [Medline: 33059959]
15. Country Progress Report - Zambia: Global AIDS Monitoring 2020. Joint United Nations Programme on HIV/AIDS. 2020. URL: https://www.unaids.org/sites/default/files/country/documents/ZMB_2020_countryreport.pdf [accessed 2021-09-26]
16. The safeguard young people programme: three years on: addressing the urgent needs of youth across Southern Africa. United Nations Population Fund. 2017. URL: https://esaro.unfpa.org/sites/default/files/pub-pdf/Phase%201%20-%20SYP%20Report_0.pdf [accessed 2021-09-26]
17. End of project evaluation of the BBC media action radio waves and Tikambe projects in Zambia: final report. Sida Decentralised Evaluation. 2019. URL: https://cdn.sida.se/publications/files/sida62268en-end-of-project-evaluation-of-the-bbc-media-action-radio-waves-and-tikambe-projects-in-zambia.pdf [accessed 2021-09-26]
18. U-Report Zambia. United Nations International Children's Emergency Fund. URL: https://www.unicef.org/zambia/u-report-zambia [accessed 2021-09-26]
19. Ministry of Health, Zambia. COVID-19 update #12 summary. Facebook. 2020. URL: https://www.Facebook.com/mohzambia/ [accessed 2021-09-26]
20. Nolan C, Packel L, Hope R, Levine J, Baringer L, Gatare E, et al. Design and impact evaluation of a digital reproductive health program in Rwanda using a cluster randomized design: study protocol. BMC Public Health 2020 Nov 13;20(1):1701 [FREE Full text] [doi: 10.1186/s12889-020-09746-7] [Medline: 33187485]
21. Castle S, Martha S. Family planning and youth in West Africa: mass media, digital media, and social and behavior change communication strategies. Breakthrough RESEARCH Literature Review. 2019 Aug. URL: https://breakthroughactionandresearch.org/wp-content/uploads/2019/09/Mass-Media-Literature-Review.pdf [accessed 2021-10-02]
22. Faust L, Yaya S. The effect of HIV educational interventions on HIV-related knowledge, condom use, and HIV incidence in sub-Saharan Africa: a systematic review and meta-analysis. BMC Public Health 2018 Nov 13;18(1):1254 [FREE Full text] [doi: 10.1186/s12889-018-6178-y] [Medline: 30424761]
23. Winskell K, Sabben G, Akelo V, Ondeng'e K, Obong'o C, Stephenson R, et al. A smartphone game-based intervention (Tumaini) to prevent HIV among young Africans: pilot randomized controlled trial. JMIR Mhealth Uhealth 2018 Aug 01;6(8):e10482 [FREE Full text] [doi: 10.2196/10482] [Medline: 30068501]
24. Qualtrics. 2018. URL: https://www.qualtrics.com/au/ [accessed 2021-05-28]
25. Kok G, Gottlieb NH, Peters GJ, Mullen PD, Parcel GS, Ruiter RA, et al. A taxonomy of behaviour change methods: an intervention mapping approach. Health Psychol Rev 2016 Sep;10(3):297-312 [FREE Full text] [doi: 10.1080/17437199.2015.1077155] [Medline: 26262912]
26. Turner JC. Social identity theory: where are we now? In: Ellemers N, Spears R, Doosje B, editors. Social Identity: Context, Commitment, Content. Hoboken, NJ, USA: Wiley-Blackwell; 1999:6-34.
27. Protogerou C, Johnson BT, Hagger MS. An integrated model of condom use in Sub-Saharan African youth: a meta-analysis. Health Psychol 2018 Jun;37(6):586-602 [FREE Full text] [doi: 10.1037/hea0000604] [Medline: 29708390]
28. Klasnja P, Hekler EB. Rethinking evaluations of mHealth systems for behavior change. GetMobile 2018 Jun;22(2):11-14 [FREE Full text] [doi: 10.1145/3276145.3276149] [Medline: 30680312]

29. DeSmet A, Shegog R, Van Ryckeghem D, Crombez G, De Bourdeaudhuij I. A systematic review and meta-analysis of interventions for sexual health promotion involving serious digital games. Games Health J 2015 Apr;4(2):78-90. [doi: 10.1089/g4h.2014.0110] [Medline: 26181801]
30. Jessen S, Mirkovic J, Ruland CM. Creating gameful design in mHealth: a participatory co-design approach. JMIR Mhealth Uhealth 2018 Dec 14;6(12):e11579 [FREE Full text] [doi: 10.2196/11579] [Medline: 30552080]
31. cGrate. 2012. URL: https://www.cgrate.com/index.php [accessed 2021-05-10]
32. Stanton B, Deveaux L, Lunn S, Yu S, Brathwaite N, Li X, et al. Condom-use skills checklist: a proxy for assessing condom-use knowledge and skills when direct observation is not possible. J Health Popul Nutr 2009 Jun;27(3):406-413 [FREE Full text] [doi: 10.3329/jhpn.v27i3.3383] [Medline: 19507756]
33. Creswell JW. Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research. 4th edition. Hoboken, NJ, USA: Pearson; 2010.
34. 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]
35. Muhammad NA, Shamsuddin K, Mohd Amin R, Omar K, Thurasamy R. Questionnaire development and validity to measure sexual intention among youth in Malaysia. BMC Public Health 2017 Feb 02;17(1):157 [FREE Full text] [doi: 10.1186/s12889-016-3949-1] [Medline: 28152993]
36. National Survey of Adolescents and Young Adults: Sexual health knowledge, attitudes and experiences. Kaiser Family Foundation. 2013. URL: https://www.kff.org/wp-content/uploads/2013/01/national-survey-of-adolescents-and-young-adults.pdf [accessed 2019-09-30]
37. Lewis IM, Watson B, White KM. Response efficacy: the key to minimizing rejection and maximizing acceptance of emotion-based anti-speeding messages. Accid Anal Prev 2010 Mar;42(2):459-467. [doi: 10.1016/j.aap.2009.09.008] [Medline: 20159067]
38. Stata – Statistical Software: Release 16.1. StatCorp. 2019. URL: https://www.stata.com [accessed 2021-08-09]
39. Wickham H, Francois R, Henry L, Müller K. dplyr: A Grammar of Data Manipulation. R package version 1.0.7. 2021. URL: https://cran.r-project.org/web/packages/dplyr/index.html [accessed 2022-01-11]
40. Watkins DC. Rapid and rigorous qualitative data analysis: the "RADaR" technique for applied research. Int J Qual Methods 2017 Jun 08;16(1):160940691771213. [doi: 10.1177/1609406917712131]
41. Garcia PJ, Miranda AE, Gupta S, Garland SM, Escobar ME, Fortenberry JD, International Union Against Sexually Transmitted Infections. The role of sexually transmitted infections (STI) prevention and control programs in reducing gender, sexual and STI-related stigma. EClinicalMedicine 2021 Mar;33:100764 [FREE Full text] [doi: 10.1016/j.eclinm.2021.100764] [Medline: 33718850]
42. Lee AS, Cody SL. The stigma of sexually transmitted infections. Nurs Clin North Am 2020 Sep;55(3):295-305. [doi: 10.1016/j.cnur.2020.05.002] [Medline: 32762851]
43. Thomas JA, Ditchman N, Beedle RB. The impact of knowledge, self-efficacy, and stigma on STI testing intention among college students. J Am Coll Health (forthcoming) 2020 Aug 13:1-11. [doi: 10.1080/07448481.2020.1799808] [Medline: 32790582]
44. Bender S, Hill K, Information Resources Management Association (USA). Sexually transmitted infections – diagnoses, stigma, and mental health. In: Research Anthology on Mental Health Stigma, Education, and Treatment. Hershey, PA, USA: IGI Global; 2021:103-120.
45. Darabi F, Yaseri M, Kaveh MH, Khalajabadi Farahani F, Majlessi F, Shojaeizadeh D. The effect of a theory of planned behavior-based educational intervention on sexual and reproductive health in Iranian adolescent girls: a randomized controlled trial. J Res Health Sci 2017 Nov 28;17(4):e00400 [FREE Full text] [Medline: 29233954]
46. Morales A, Espada JP, Orgilés M, Escribano S, Johnson BT, Lightfoot M. Interventions to reduce risk for sexually transmitted infections in adolescents: a meta-analysis of trials, 2008-2016. PLoS One 2018 Jun 28;13(6):e0199421 [FREE Full text] [doi: 10.1371/journal.pone.0199421] [Medline: 29953546]
47. Quilling R, Pieterse G. To see someone else's perspective: a case for digital stories in schools. J Indep Teach Learn 2018 Jun 12;13(1):41-56 [FREE Full text]
48. Barrett M, Laris BA, Anderson P, Baumler E, Gerber A, Kesler K, et al. Condom use and error experience among young adolescents: implications for classroom instruction. Health Promot Pract 2021 May;22(3):313-317. [doi: 10.1177/1524839920935431] [Medline: 32536213]
49. Levkoff L, Kempner M. We're teaching about condoms all wrong: how sex educators reinforce negative attitudes and misinformation about condoms and how to change that. Am J Sex Educ 2021 Aug 09;16(4):466-479. [doi: 10.1080/15546128.2021.1960664]
50. Chatterjee K. What can we learn about the female condom online? An analysis of visual representations of the female condom on the Internet. Atl J Commun 2018 Jun;26(3):149-163. [doi: 10.1080/15456870.2018.1474217]
51. Penal Code Section 177(1) Obscene matter or things in Chapter 87 of the Laws of Zambia. Laws of Zambia, Republic of Zambia. URL: https://www.parliament.gov.zm/sites/default/files/documents/acts/Penal%20Code%20Act.pdf [accessed 2021-11-08]

52. Section 102 of the Electronic Communications and Transactions ACT of Zambia. United Nations Office on Drugs and Crime. URL: https://sherloc.unodc.org/cld/en/legislation/zmb/electronic_communications_and_transactions_act/part_xv/section_102/section_102.html? [accessed 2021-11-08]
53. Goodyear V, Armour K, Wood H. The Impact of Social Media on Young People's Health and Wellbeing: Evidence, Guidelines and Actions. University of Birmingham. 2018 Jan. URL: http://epapers.bham.ac.uk/3070/1/The_Impact_of_Social_Media_on_Young_People%E2%80%99s_Health_and_Wellbeing_(FINAL_15TH_JAN).pdf [accessed 2021-11-10]
54. Closson K, Dietrich JJ, Lachowsky NJ, Nkala B, Palmer A, Cui Z, et al. Sexual self-efficacy and gender: a review of condom use and sexual negotiation among young men and women in Sub-Saharan Africa. J Sex Res 2018;55(4-5):522-539. [doi: 10.1080/00224499.2017.1421607] [Medline: 29466024]
55. Cislaghi B, Shakya H. Social norms and adolescents' sexual health: an introduction for practitioners working in low and mid-income African countries. Afr J Reprod Health 2018 Mar;22(1):38-46 [FREE Full text] [doi: 10.29063/ajrh2018/v22i1.4] [Medline: 29777641]
56. Wood JM, Harries J, Kalichman M, Kalichman S, Nkoko K, Mathews C. Exploring motivation to notify and barriers to partner notification of sexually transmitted infections in South Africa: a qualitative study. BMC Public Health 2018 Aug 06;18(1):980 [FREE Full text] [doi: 10.1186/s12889-018-5909-4] [Medline: 30081960]
57. Nyblade L, Stockton M, Nyato D, Wamoyi J. Perceived, anticipated and experienced stigma: exploring manifestations and implications for young people's sexual and reproductive health and access to care in North-Western Tanzania. Cult Health Sex 2017 Oct;19(10):1092-1107. [doi: 10.1080/13691058.2017.1293844] [Medline: 28276918]
58. Chitneni P, Beksinska M, Dietrich JJ, Jaggernath M, Closson K, Smith P, AYAZAZI Research Team. Partner notification and treatment outcomes among South African adolescents and young adults diagnosed with a sexually transmitted infection via laboratory-based screening. Int J STD AIDS 2020 Jun;31(7):627-636 [FREE Full text] [doi: 10.1177/0956462420915395] [Medline: 32403988]
59. Nmadu AG, Mohamed S, Usman NO. Adolescents' utilization of reproductive health services in Kaduna, Nigeria: the role of stigma. Vulnerable Child Youth Stud 2020 Jul 31;15(3):246-256. [doi: 10.1080/17450128.2020.1800156]
60. Bylund S, Målqvist M, Peter N, Herzig van Wees S. Negotiating social norms, the legacy of vertical health initiatives and contradicting health policies: a qualitative study of health professionals' perceptions and attitudes of providing adolescent sexual and reproductive health care in Arusha and Kilimanjaro region, Tanzania. Glob Health Action 2020 Dec 31;13(1):1775992 [FREE Full text] [doi: 10.1080/16549716.2020.1775992] [Medline: 32588782]
61. Hager B, Hasselberg M, Arzubi E, Betlinski J, Duncan M, Richman J, et al. Leveraging behavioral health expertise: practices and potential of the project ECHO approach to virtually integrating care in underserved areas. Psychiatr Serv 2018 Apr 01;69(4):366-369. [doi: 10.1176/appi.ps.201700211] [Medline: 29446334]
62. Manlove J, Whitfield B, Finocharo J, Cook E. Lessons learned from replicating a randomized control trial evaluation of an app-based sexual health program. Int J Environ Res Public Health 2021 Mar 23;18(6):3305 [FREE Full text] [doi: 10.3390/ijerph18063305] [Medline: 33806809]
63. Scheinfeld E. Shame and STIs: an exploration of emerging adult students' felt shame and stigma towards getting tested for and disclosing sexually transmitted infections. Int J Environ Res Public Health 2021 Jul 05;18(13):7179 [FREE Full text] [doi: 10.3390/ijerph18137179] [Medline: 34281116]
64. Fumpa-Makano R. Girls' initiation ceremonies in Zambia: reflections on their role in girl child educational advancement. Int J Arts Soc Sci 2019;2(4):22-34 [FREE Full text]
65. Kok MC, van Eldik Z, Kakal T, Munthali A, Menon JA, Pires P, et al. Being dragged into adulthood? Young people's agency concerning sex, relationships and marriage in Malawi, Mozambique and Zambia. Cult Health Sex 2022 Jun;24(6):767-781. [doi: 10.1080/13691058.2021.1881618] [Medline: 33630727]
66. Widman L, Evans R, Javidi H, Choukas-Bradley S. Assessment of parent-based interventions for adolescent sexual health: a systematic review and meta-analysis. JAMA Pediatr 2019 Sep 01;173(9):866-877 [FREE Full text] [doi: 10.1001/jamapediatrics.2019.2324] [Medline: 31355860]
67. Wamoyi J, Fenwick A, Urassa M, Zaba B, Stones W. Parent-child communication about sexual and reproductive health in rural Tanzania: implications for young people's sexual health interventions. Reprod Health 2010 May 12;7:6 [FREE Full text] [doi: 10.1186/1742-4755-7-6] [Medline: 20462413]
68. Isaksen KJ, Musonda P, Sandøy IF. Parent-child communication about sexual issues in Zambia: a cross sectional study of adolescent girls and their parents. BMC Public Health 2020 Jul 16;20(1):1120 [FREE Full text] [doi: 10.1186/s12889-020-09218-y] [Medline: 32677930]
69. Nyimbili F, Mainza R, Mumba L, Katunansa B. Teacher and parental involvement in providing comprehensive sexuality education in selected primary schools of Kalomo district of Zambia. J Adult Educ 2019 Dec 5;1(2):53-62 [FREE Full text]
70. D'Cruz J, Santa Maria D, Dube S, Markham C, McLaughlin J, Wilkerson JM, et al. Promoting parent-child sexual health dialogue with an intergenerational game: parent and youth perspectives. Games Health J 2015 Apr;4(2):113-122 [FREE Full text] [doi: 10.1089/g4h.2014.0080] [Medline: 26181805]
71. Kemp S. Digital 2021: Zambia. Datareportal. 2021 Feb 12. URL: https://datareportal.com/reports/digital-2021-zambia [accessed 2021-11-09]

72. Simungala G, Jimaima H. Multilingual realities of language contact at the University of Zambia. J Asian Afr Stud 2020 Dec 15;56(7):1644-1657. [doi: 10.1177/0021909620980038]
73. Barnes BR. Quasi-experimental designs in applied behavioural health research. In: Laher S, Fynn A, Kramer S, editors. Transforming Research Methods in the Social Sciences: Case Studies from South Africa. Witwatersrand, South Africa: Wits University Press; 2019:84-96.
74. Ybarra ML, Korchmaros JD, Prescott TL, Birungi R. A randomized controlled trial to increase HIV preventive information, motivation, and behavioral skills in Ugandan adolescents. Ann Behav Med 2015 Jun;49(3):473-485 [FREE Full text] [doi: 10.1007/s12160-014-9673-0] [Medline: 25633626]
75. Haruna H, Hu X, Chu SK, Mellecker RR, Gabriel G, Ndekao PS. Improving sexual health education programs for adolescent students through game-based learning and gamification. Int J Environ Res Public Health 2018 Sep 17;15(9):2027 [FREE Full text] [doi: 10.3390/ijerph15092027] [Medline: 30227642]
76. Bull S, Pratte K, Whitesell N, Rietmeijer C, McFarlane M. Effects of an Internet-based intervention for HIV prevention: the Youthnet trials. AIDS Behav 2009 Jun;13(3):474-487 [FREE Full text] [doi: 10.1007/s10461-008-9487-9] [Medline: 19037719]
77. Mevissen FE, Ruiter RA, Meertens RM, Zimbile F, Schaalma HP. Justify your love: testing an online STI-risk communication intervention designed to promote condom use and STI-testing. Psychol Health 2011 Feb;26(2):205-221. [doi: 10.1080/08870446.2011.531575] [Medline: 21318930]
78. Ybarra ML, Bull SS, Prescott TL, Korchmaros JD, Bangsberg DR, Kiwanuka JP. Adolescent abstinence and unprotected sex in CyberSenga, an Internet-based HIV prevention program: randomized clinical trial of efficacy. PLoS One 2013 Aug 14;8(8):e70083 [FREE Full text] [doi: 10.1371/journal.pone.0070083] [Medline: 23967069]

Abbreviations

aATE: adjusted average treatment effect
aOR: adjusted odds ratio
BITKZ: Be in the know Zambia
DHI: digital health intervention
IDI: in-depth interview
SRH: sexual and reproductive health
SSA: sub-Saharan Africa
STI: sexually transmitted infection
ZMW: Zambian Kwacha

Edited by R Kukafka; submitted 27.02.22; peer-reviewed by G Sabben, S Skeen; comments to author 02.04.22; revised version received 14.04.22; accepted 29.05.22; published 07.07.22.

Please cite as:
Sharma A, Mwamba C, Ng'andu M, Kamanga V, Zoonadi Mendamenda M, Azgad Y, Jabbie Z, Chipungu J, Pry JM
Pilot Implementation of a User-Driven, Web-Based Application Designed to Improve Sexual Health Knowledge and Communication Among Young Zambians: Mixed Methods Study
J Med Internet Res 2022;24(7):e37600
URL: https://www.jmir.org/2022/7/e37600
doi:10.2196/37600
PMID:35797099

Original Paper

Dosage Frequency Effects on Treatment Outcomes Following Self-managed Digital Therapy: Retrospective Cohort Study

Claire Cordella[1], PhD; Michael Munsell[2], PhD; Jason Godlove[2], PhD; Veera Anantha[2], PhD; Mahendra Advani[2], MS; Swathi Kiran[1,2], PhD

[1]Department of Speech, Language and Hearing Sciences, Boston University, Boston, MA, United States

[2]Constant Therapy Health, Lexington, MA, United States

Corresponding Author:
Claire Cordella, PhD
Department of Speech
Language and Hearing Sciences
Boston University
635 Commonwealth Ave
Boston, MA, 02215
United States
Phone: 1 617 353 2706
Email: cordella@bu.edu

Abstract

Background: Although the efficacy of high-dose speech-language therapy (SLT) for individuals with poststroke aphasia has been established in the literature, there is a gap in translating these research findings to clinical practice. Therefore, patients continue to receive suboptimal amounts of SLT, with negative consequences for their functional communication recovery. Recent research has identified self-managed digital health technology as one way to close the dosage gap by enabling high-intensity therapy unrestricted by clinician availability or other practical constraints. However, there is limited empirical evidence available to rehabilitation professionals to guide dose prescriptions for self-managed SLT despite their increasing use in the COVID-19 era and likely beyond.

Objective: This study aims to leverage real-world mobile health data to investigate the effects of varied dosage frequency on performance outcomes for individuals with poststroke speech, language, and cognitive deficits following a 10-week period of self-managed treatment via a commercially available digital health platform.

Methods: Anonymized data from 2249 poststroke survivors who used the Constant Therapy app between late 2016 and 2019 were analyzed. The data included therapy tasks spanning 13 different language and cognitive skill domains. For each patient, the weekly therapy dosage was calculated based on the median number of days per week of app use over the 10-week therapy period, binned into groups of 1, 2, 3, 4, or ≥5 days per week. Linear mixed-effects models were run to examine change in performance over time as a function of dosage group, with post hoc comparisons of slopes to evaluate the performance gain associated with each additional day of practice.

Results: Across all skill domains, linear mixed-effects model results showed that performance improvement was significantly greater for patients who practiced 2 (β=.001; $t_{15,355}$=2.37; P=.02), 3 (β=.003; t_{9738}=5.21; P<.001), 4 (β=.005; t_{9289}=7.82; P<.001), or ≥5 (β=.005; t_{6343}=8.14; P<.001) days per week compared with those who only practiced for 1 day per week. Post hoc comparisons confirmed an incremental dosage effect accumulating with each day of practice (ie, 1 day vs 2 days, 2 days vs 3 days, and 3 days vs 4 days), apart from 4 days versus ≥5 days of practice per week. The result of greater improvement for higher versus lower dosage frequency groups was true not only across all domains but also within a majority of individual subdomains.

Conclusions: The findings from this study demonstrated that increased dosage frequency is associated with greater therapy gains over a 10-week treatment period of self-managed digital therapy. The use of real-world data maximizes the ecological validity of study results and makes the findings more generalizable to clinical settings. This study represents an important step toward the development of optimal dose recommendations for self-managed SLT.

(J Med Internet Res 2022;24(7):e36135) doi:10.2196/36135

KEYWORDS

aphasia; stroke; technology; rehabilitation; dosage

Introduction

Background

Approximately one-third of all strokes result in aphasia or other communication disorders that affect a person's ability to speak, understand, read or write [1]. For a significant number of stroke survivors with aphasia—an estimated 2.25 million in the United States and the United Kingdom [1]—these communication deficits portend poorer global health outcomes (compared with stroke survivors without aphasia), including higher overall mortality, reduced functional recovery, social isolation, and reduced overall quality of life [2-5].

Fortunately, speech-language therapy (SLT) is an effective means for improving impairment- and participation-based language outcomes in individuals with chronic poststroke aphasia. A comprehensive Cochrane review of SLT randomized control trials reported greater benefits to communication when patients with chronic aphasia received therapy at high intensity (from 4 to 15 hours per week), high dosage (27-208 hours in total), or over a long period (up to 22 months) compared to more moderate treatment schedules [6].

Despite the evidence that supports the provision of high-dose SLT to stroke survivors with aphasia, patients are often unable to access sufficiently intense therapy as part of their usual care. Across the English-speaking world, it is estimated that individuals with chronic poststroke aphasia receive, on average, <5 hours of therapy per week [7,8], far less than the recommended 5 to 10 hours per week that is typical of evidence-based intensive therapy regimens [6,9]. This reality of insufficient usual care is caused by several barriers that practically limit patients' access to SLT, such as provider shortages, caps on Medicare reimbursement, geographic isolation, and lack of transportation, among other factors [10,11].

One way to offset the lack of sufficient therapy is to enable patients to engage in in-home practice through computerized or app-based therapeutic programs [11,12]. Many studies have evaluated digital SLT interventions as part of a treatment protocol, delivered as tablet- or computer-based programs [13-26]. A smaller subset of studies have investigated self-managed programs, in which users not only complete therapy at home but also determine their own practice schedule [22-26]. Crucially, the freedom to determine one's own practice schedule means that dose parameters for these types of therapies can and do vary widely from patient to patient [27]. This naturally occurring variance in dosage presents a unique opportunity to probe dose-response relationships in SLT. Dose articulation studies are a critical first step toward establishing optimal dosage recommendations for SLT interventions [28,29]. To date, only a handful of studies have directly compared different dosage amounts of the same intervention, and none have done so in the context of self-managed digital therapies [30-36].

Objective

In this study, we leveraged real-world mobile health data to investigate the effects of different dose levels on treatment outcomes following a 10-week treatment period using a commercially available digital health platform. Noting that a consensus on the definition of SLT dosage and intensity has not been definitively reached in the literature [28,29,37-39], we chose to focus on dosage frequency, which is defined as the number of days per week during which a patient completes a therapy session. This measure is easily generalizable across patients and applicable to clinical settings. In this analysis, we retrospectively examined how often users completed computer-based therapy sessions with the Constant Therapy program and evaluated the relationship between their dosage frequency and improvement over the treatment period in several functional domains. It was hypothesized that patients who adhered to a greater dosage frequency of therapy would see greater improvement during the first 10 weeks of treatment than individuals at the lowest dosage frequency.

Methods

Participants

Data were aggregated and analyzed from patients who used the Constant Therapy app between October 2016 and October 2019. The data were anonymized before being shared for analysis with Boston University. All users (N=238,767) consented to the use of their exercise and therapy performance data for research purposes. Constant Therapy users were asked to provide basic demographic and diagnostic information upon initial sign-up, including age, time since injury, sex, and diagnoses (eg, stroke, aphasia, and traumatic brain injury). For this study, only users who reported having had a stroke with resultant speech, language, and cognitive deficits were included for analysis. An additional inclusion criterion was applied that required users to engage with the app for at least one day in 10 of their first 15 calendar weeks of use. The resultant study sample included 2249 unique patients with speech, language, and cognitive deficits following stroke. Across the entire sample and within each dosage group, the most commonly endorsed diagnoses were stroke alone; stroke and aphasia; and stroke, aphasia, and apraxia. Dosage groups were determined by first calculating, per individual, the median number of days per week of Constant Therapy use over the 10-week therapy period of interest, and then binning into categories of 1, 2, 3, 4, or ≥5 days per week.

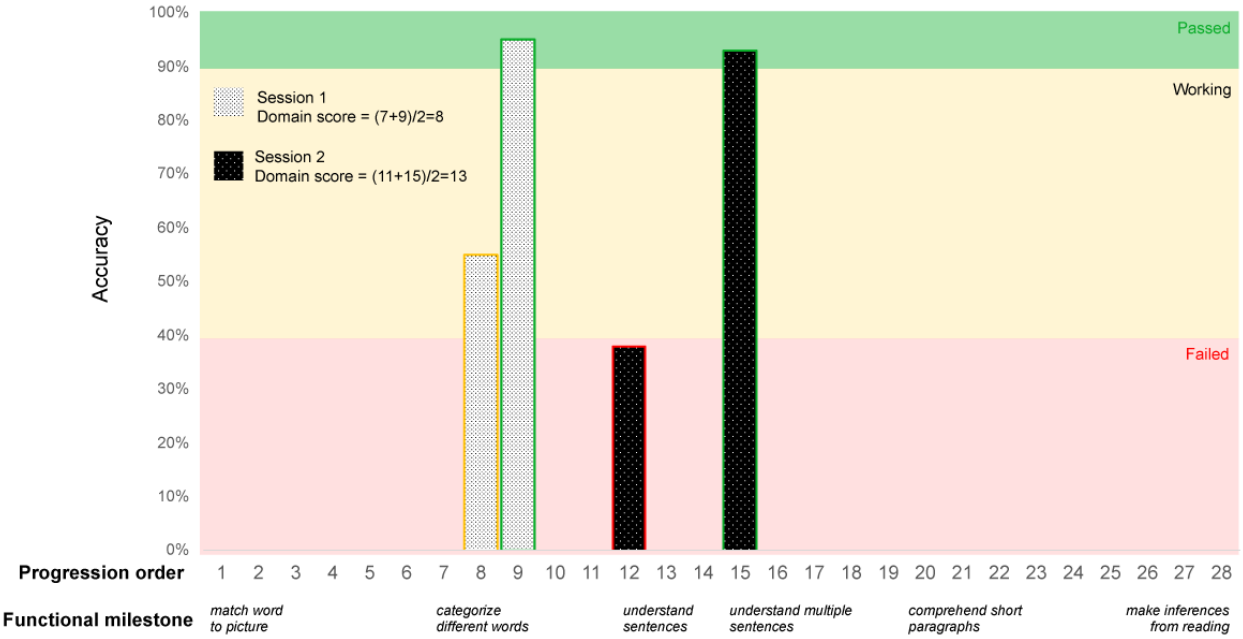
Therapy Program

Constant Therapy [40] is an evidence-based digital therapeutic that features over 244 individual tasks spanning various speech, language, and cognitive skill domains [22,25,26,41,42]. This study focused on task data for the following 13 domains: (1) auditory comprehension, (2) phonological processing, (3) production, (4) reading, (5) writing, (6) naming, (7) attention, (8) auditory memory, (9) visual memory, (10) analytical, (11) arithmetic, (12) quantitative, and (13) visuospatial skills.

Importantly, users can tailor their therapy program by self-selecting the skill domains in need of improvement, meaning that the specific tasks being worked on as part of therapy differ from user to user. Task difficulty is also adjusted on a user-by-user basis, based on an adaptive difficulty algorithm that advances users to a more difficult version of a given task once they have achieved mastery. During a session, patients practice tasks in order of increasing level of difficulty. The order in which subsequent, more difficult tasks are assigned is determined by a universal task progression order per domain, whereby every task type is ranked serially from least to most difficult. The progression order for each skill domain was structured based on research evidence and clinician consultation and fine-tuned using population performance [43].

Within each practice session, each task is practiced until accuracy reaches 90% or higher on ≥2 occasions, at which point a patient is advanced to the next level of difficulty or to a different task. In addition, if a user was not improving on a level or their accuracy was below 40%, a lower level of the task was assigned in addition to or in replacement of the original task. The Constant Therapy program records task performance data (ie, accuracy) and session activities, including usability logs, time stamps, and item completion indicators. On the basis of these performance data, a domain score is calculated to provide a summative assessment of a user's performance in a specific skill domain, considering that users are completing tasks at various difficulty levels. Generation of the domain score involves (1) identifying the highest task passed (accuracy ≥90%) or working (accuracy between 40% and 90%) and the lowest task working or failed (accuracy <40%) during a session and (2) taking the average progression order of the highest level passed or working and the lowest level working or failed, thereby providing an estimate of the given session's difficulty level. The progression order for failed or working tasks is adjusted by subtracting 1, because the highest difficulty level successfully passed at that time is represented by the previous task in the progression order. Scores are normalized by dividing by the total number of task levels—which varies by domain—to make scores comparable across different skill domains. Therefore, change in domain score can be interpreted as a patient's increase in difficulty level as a percentage of a domain's total items. An example domain score calculation for a hypothetical patient in the reading domain is provided in Figure 1. Domain scores were averaged across sessions if multiple sessions occurred in a single week.

Figure 1. Example calculated domain score (reading domain). Tasks are introduced to a patient according to a domain's progression order (x-axis, top row). The level of function a patient should be able to demonstrate after successfully passing the listed task is noted as the functional milestone (x-axis, bottom row). Shaded bars show the task accuracy scores for a hypothetical patient's highest task passed and lowest task working or failed across 2 different sessions.



This study used the calculated weekly domain scores for 2 purposes. First, weekly domain scores served as the dependent variable of interest to index performance change over the intervention period. The weekly domain score was the average domain score across all sessions completed by the user per week, calculated for each of the 10 weeks of the intervention period. Second, we extracted the domain score at baseline (ie, week 0 of the intervention period) as a means by which to index initial severity, as standardized measures of baseline language and cognitive function were not available in the data set, given the real-world nature of the data obtained.

Statistical Analysis

For each of the 13 domains, the first week of therapy was treated as the baseline (assigned as week 0), and weekly domain scores were extracted for each of the 10 weeks of the intervention period, as described earlier. In order to examine changes in weekly domain scores over time as a function of dosage frequency group, linear mixed-effects models (LMMs) were run first for scores combined across all domains and then independently for each domain. For the overall model encompassing all domains, the weekly domain score served as the dependent variable, with fixed effects of time (week

number), dosage frequency group, cumulative practice amount (ie, total hours spent completing therapy tasks), time × dosage frequency group, and time × cumulative practice amount. Covariates of age, time since stroke (≤6 and >6 months), sex, and baseline domain scores were also included as fixed effects in the model. The model included random effects of patients and domains. This final model structure was determined through an iterative process of stepwise addition of model terms, beginning with the determination of the optimal random effects structure and proceeding to the determination of the optimal fixed-effects structure. Nested models were compared using Likelihood-Ratio Tests with additional reference to the Akaike information criterion and Bayesian information criterion values of each candidate model. The LMM building and selection process is reported in detail in Multimedia Appendix 1, following best-practice conventions for LMM reporting in psychological science [44]. This same model structure was applied to the analyses of each of the 13 individual domains, except that for these analyses, the random effect of the domain was excluded. All statistical analyses were conducted in R (version 4.0.2; R Foundation for Statistical Computing) using *lme4*, *lmerTest*, *emmeans*, and *sjPlot* packages [45-49].

Ethics Approval

This project was considered an institutional review board–exempt retrospective analysis by Pearl Institutional Review Board (#17-LNCO-101) under 45 Code of Federal Regulations 46.101(b) category 2.

Results

All Skill Domains

Data of 2249 patients with poststroke deficits in speech, language, or cognitive deficits were analyzed in this study. The average age of the sample was 63 (SD 14) years, and the majority of patients (N=1319) were in the acute recovery stage (ie, ≤6 months poststroke). The average (normalized) baseline domain score was 33% (SD 20%), indicating that the patients were typically in the lower third of the domain's task progression order during their first week of therapy. The dosage groups did not significantly differ in terms of age, sex, or proportion of patients with acute condition. With regard to age, digital literacy did not appear to be a barrier to use among older adults in the sample, as older users showed similar practice patterns as younger users, in line with previously published findings showing robust engagement with the Constant Therapy app among older users [27]. Significant overall differences across dosage frequency groups were observed for baseline domain score ($F_{4,57898}$=6.937; P<.001) and total hours of therapy ($F_{4,61197}$=54.54; P<.001), although effect sizes between dosage groups were uniformly small for both measures (Multimedia Appendix 1). Nonetheless, these factors were included as covariates in all analysis models to account for the potential confounding effects of severity (ie, baseline domain score) and cumulative therapy exposure (ie, total hours of therapy). The summary statistics for the entire cohort and for each dosage frequency group are presented in Table 1.

Table 1. Summary statistics of study cohort (N=2249).

Characteristics	Overall (N=2249)	By dosage frequency group				
		1 day per week (N=888)	2 days per week (N=1155)	3 days per week (N=804)	4 days per week (N=574)	5 days per week (N=481)
Age (years), mean (SD)	63 (14)	64 (14)	64 (14)	63 (13)	63 (13)	63 (13)
Sex, n (%)						
Male	1269 (56.4)	500 (56.3)	645 (55.8)	459 (57.1)	335 (58.4)	277 (57.6)
Female	968 (43)	384 (43.2)	506 (43.8)	343 (42.7)	236 (41.1)	199 (41.4)
Not specified	12 (0.5)	4 (0.5)	4 (0.3)	2 (0.2)	3 (0.5)	5 (1)
Chronicity, n (%)						
Acute (≤6 months)	1319 (58.6)	494 (55.6)	671 (58.1)	463 (57.6)	335 (58.4)	294 (61.1)
Chronic (>6 months)	930 (41.4)	394 (44.4)	484 (41.9)	341 (42.4)	239 (41.6)	187 (38.9)
Baseline domain score[a], mean (SD)	0.33 (0.20)	0.33 (0.21)	0.33 (0.20)	0.34 (0.20)	0.34 (0.20)	0.34 (0.20)
Total hours[a], mean (SD)	6.2 (22)	3.7 (21.4)	5.3 (28.6)	6.1 (22.8)	6.9 (11.4)	10.6 (11.9)

[a]Baseline domain score and total hours variables are calculated per individual skill domain.

Across all skill domains, the model results (Tables 2 and 3) revealed significant main effects of time ($F_{1,15}$=106.46; P<.001), time since stroke ($F_{1,1753}$=16.57; P<.001), baseline domain score ($F_{1,104,365}$=67,301.21; P<.001), cumulative practice amount ($F_{1,60502}$=12.83; P<.001), and dosage group frequency ($F_{4,8873}$=6.22; P<.001) on domain score in the 10-week treatment period. Specifically, a greater weekly domain score was associated with an increase in the number of weeks of therapy (β=.009; t=−12.27; P<.001), acute condition (β=.010; t=4.07; P<.001), higher baseline domain score (β=.662; t=259.42; P<.001), greater cumulative practice amount (β=.0001; t=3.58; P<.001), and greater practice frequency (2 days: β=.001, t=0.52, P=.60; 3 days: β=.008, t=3.16, P=.002; 4 days: β=.008, t=2.91, P=.004; ≥5 days: β=.011, t=3.95, P<.001). Age and sex were not significant predictors of domain score, nor was the interaction of time × cumulative practice amount.

Crucial to our question of interest, the time × dosage frequency group interaction was significant ($F_{4,10347}$=6.22; P<.001), indicating that although we see gains in domain score for all dosage groups over time (Figure 2A), the rate of improvement is highly dependent on the frequency of practice. Rates of improvement were significantly greater for patients who practiced 2 (β=.001; $t_{15,355}$ =2.37; P=.02), 3 (β=.003; t_{9738}=5.21; P<.001), 4 (β=.005; t_{9289}=7.82; P<.001), or ≥5 (β=.005; t_{6343}=8.14; P<.001) days per week than for those who only practiced 1 day per week. Furthermore, post hoc pairwise comparison of slopes (Table 4) showed an incremental dosage effect accumulating with each additional day of practice (ie, 1 day vs 2 days, 2 days vs 3 days, and 3 days vs 4 days), apart from 4 days versus ≥5 days a week of practice. Table 5 presents a pairwise comparison of estimated means per dosage frequency group at the beginning (ie, week 0) and end (ie, week 9) of treatment, illustrating that although at baseline, domain scores between incremental dosage groups (ie, 1 day vs 2 days, 2 days vs 3 days, 3 days vs 4 days, and 4 days vs ≥5 days) were not significant, by the end of treatment, significant differences in means emerged for all group comparisons except for the groups practicing 4 days versus ≥5 days per week. This result indicates that the significant magnitude differences accrued over the course of treatment are attributable to differences in slopes across the dosage groups as opposed to baseline differences in means. Figure 2B shows the cumulative effect of treatment—calculated as the standardized pretreatment versus posttreatment effect size per dosage group—and demonstrates that although there was at least a moderate treatment effect for all dosage groups, this effect was larger for patients who practiced more frequently. The standardized effect size was calculated for each dosage frequency group based on the difference in LMM-generated estimated marginal means from pretreatment (ie, week 0) to posttreatment (ie, week 9), using the eff_size function in the *emmeans* package in R.

Table 2. Final linear mixed-effects model results summary (fixed effects), across all skill domains[a,b].

Predictors	Estimates (SE)	*t* test (*df*)	*P* value
Fixed effects			
Intercept	1.36×10^{-1} (1.18×10^{-2})	12.27 (2.51×10^{1})	*<.001*[c]
Week	9.28×10^{-3} (1.22×10^{-3})	7.62 (1.73×10^{1})	*<.001*
Dosage group (2 days per week)	1.15×10^{-3} (2.22×10^{-3})	0.52 (1.34×10^{4})	.60
Dosage group (3 days per week)	8.00×10^{-3} (2.54×10^{-3})	3.16 (8.21×10^{3})	.002
Dosage group (4 days per week)	8.30×10^{-3} (2.85×10^{-3})	2.91 (7.91×10^{3})	.004
Dosage group (≥5 days per week)	1.13×10^{-2} (2.86×10^{-3})	3.95 (5.09×10^{3})	*<.001*
Total hours	1.13×10^{-4} (3.15×10^{-5})	3.58 (6.05×10^{4})	*<.001*
Domain score baseline	6.62×10^{-1} (2.55×10^{-3})	259.42 (1.04×10^{5})	*<.001*
Age (years)	-1.54×10^{-4} (8.96×10^{-5})	−1.72 (1.79×10^{3})	.09
Sex (male)	-1.39×10^{-4} (2.43×10^{-3})	−0.06 (1.78×10^{3})	.95
Sex (not specified)	2.13×10^{-2} (1.66×10^{-2})	1.28 (1.83×10^{3})	.20
Chronicity (acute)	9.93×10^{-3} (2.44×10^{-3})	−4.07 (1.75×10^{3})	*<.001*
Week × dosage group (2 days per week)	1.13×10^{-3} (4.78×10^{-4})	2.37 (1.54×10^{4})	.02
Week × dosage group (3 days per week)	2.82×10^{-3} (5.42×10^{-4})	5.21 (9.74×10^{3})	*<.001*
Week × dosage group (4 days per week)	4.73×10^{-3} (6.05×10^{-4})	7.82 (9.29×10^{3})	*<.001*
Week × dosage group (≥5 days per week)	5.03×10^{-3} (6.17×10^{-4})	8.14 (6.34×10^{3})	*<.001*
Week × total hours	6.01×10^{-6} (4.78×10^{-6})	0.97 (6.50×10^{4})	.33

[a]N (total observations)=111,768; N (patients)=2249; N (domains)=13.

[b]Model equation: domain score (weekly average) ~ week × (dosage group + total hours) + baseline domain score + age + sex + chronicity + (1+ week:patient) + (1+ week:domain).

[c]Italicized text indicates a significant predictor, P<.001.

Table 3. Final linear mixed-effects model results summary (random effects), across all skill domains[a,b].

Predictors	Variance (SD)	Correlation
Random effects		
Residual	1.4×10^{-2} (1.2×10^{-1})	N/A[c]
Patient (intercept)	2.2×10^{-3} (4.7×10^{-2})	N/A
Domain (intercept)	1.3×10^{-3} (3.6×10^{-2})	N/A
Week:patient (slope)	1.5×10^{-5} (1.2×10^{-2})	5.2×10^{-1}
Week:domain (slope)	1.6×10^{-5} (4.0×10^{-3})	3.0×10^{-2}

[a]N (total observations)=111,768; N (patients)=2249; N (domains)=13.

[b]Model equation: domain score (weekly average) ~ week × (dosage group + total hours) + baseline domain score + age + sex + chronicity + (1+ week:patient) + (1+ week:domain).

[c]N/A: not applicable.

Figure 2. Change in domain score as a function of dosage frequency group, across all skill domains. (A) The average weekly domain score improved over the treatment period for all dosage frequency groups, but the rate of improvement was significantly greater for the higher versus lower dosage groups. Numbers in parentheses in the legend correspond to the number of unique patients in each dosage frequency group. Error bars represent the SE of the mean. (B) The treatment effect sizes were greater for the higher versus lower dosage groups. DFG: dosage frequency group (1 day per week, 2 days per week, 3 days per week, 4 days per week, ≥5 days per week).

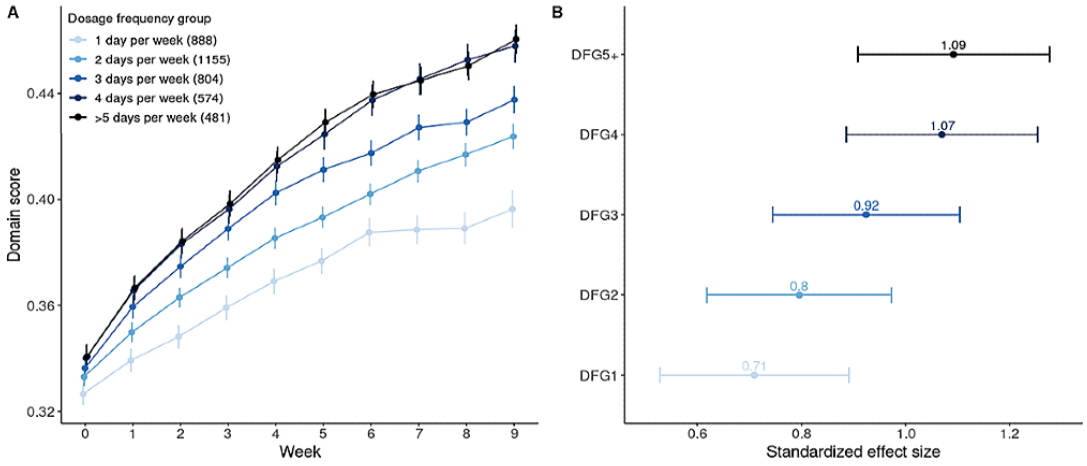


Table 4. Pairwise comparisons of slopes by dosage frequency group[a].

Contrast	Estimate (SE)	*t* test (*df*[b])	*P* value[c]
1 day per week			
2 days per week	-1.13×10^{-3} (4.77×10^{-4})	−2.37 (infinity)	.12
3 days per week	-2.82×10^{-3} (5.42×10^{-4})	−5.21 (infinity)	*<.001* [d]
4 days per week	-4.73×10^{-3} (6.05×10^{-4})	−7.82 (infinity)	*<.001*
5 days per week	-5.03×10^{-3} (6.17×10^{-4})	−8.14 (infinity)	*<.001*
2 days per week			
3 days per week	-1.69×10^{-3} (4.46×10^{-4})	−3.79 (infinity)	*.001*
4 days per week	-3.60×10^{-3} (5.30×10^{-4})	−6.79 (infinity)	*<.001*
5 days per week	-3.90×10^{-3} (5.49×10^{-4})	−7.10 (infinity)	*<.001*
3 days per week			
4 days per week	-1.91×10^{-3} (5.16×10^{-4})	−3.70 (infinity)	*.002*
5 days per week	-2.20×10^{-3} (5.56×10^{-4})	−3.97 (infinity)	*.001*
4 days per week			
5 days per week	-2.94×10^{-4} (5.56×10^{-4})	−0.52 (infinity)	.99

[a]Results are averaged over the levels of sex and chronicity.

[b]Degrees-of-freedom method: asymptotic.

[c]*P* value adjustment: Tukey method for comparing a family of 5 estimates.

[d]Italicized text indicates significant contrast, $P<.05$.

Table 5. Pairwise comparison of estimated marginal means at beginning and end of treatment.

Contrast	Estimate (SE)	*t* test (*df*)	*P* value
Week 0 (baseline)			
1 day per week			
2 days per week	-1.15×10^{-3} (2.22×10^{-3})	−0.52 (infinity)	.99
3 days per week	-8.00×10^{-3} (2.53×10^{-3})	−3.16 (infinity)	*.01*[a]
4 days per week	-8.30×10^{-3} (2.85×10^{-3})	−2.91 (infinity)	*.03*
5 days per week	-1.13×10^{-2} (2.86×10^{-3})	−3.95 (infinity)	*.001*
2 days per week			
3 days per week	-6.85×10^{-3} (2.20×10^{-3})	−3.12 (infinity)	*.02*
4 days per week	-7.15×10^{-3} (2.60×10^{-3})	−2.75 (infinity)	*.047*
5 days per week	-1.01×10^{-2} (2.63×10^{-3})	−3.86 (infinity)	*.001*
3 days per week			
4 days per week	-2.95×10^{-4} (2.61×10^{-3})	−0.11 (infinity)	.99
5 days per week	-3.30×10^{-3} (2.73×10^{-3})	−1.21 (infinity)	.75
4 days per week			
5 days per week	-3.00×10^{-3} (2.85×10^{-3})	−1.05 (infinity)	.83
Week 9 (end of analysis period)			
1 day per week			
2 days per week	-1.13×10^{-2} (3.18×10^{-3})	−3.57 (infinity)	*.003*
3 days per week	-3.34×10^{-2} (3.78×10^{-3})	−8.84 (infinity)	*<.001*
4 days per week	-5.09×10^{-2} (4.27×10^{-3})	−11.92 (infinity)	*<.001*
5 days per week	-5.65×10^{-2} (4.63×10^{-3})	−12.20 (infinity)	*<.001*
2 days per week			
3 days per week	-2.21×10^{-2} (2.85×10^{-3})	−7.74 (infinity)	*<.001*
4 days per week	-3.95×10^{-2} (3.54×10^{-3})	−11.16 (infinity)	*<.001*
5 days per week	-4.52×10^{-2} (4.00×10^{-3})	−11.28 (infinity)	*<.001*
3 days per week			
4 days per week	-1.75×10^{-2} (3.18×10^{-3})	−5.49 (infinity)	*<.001*
5 days per week	-2.31×10^{-2} (3.84×10^{-3})	−6.02 (infinity)	*<.001*
4 days per week			
5 days per week	-5.65×10^{-3} (3.67×10^{-3})	−1.54 (infinity)	.54

[a]Italicized text indicates significant contrast, $P<.05$.

Individual Skill Domains

Within individual skill domains, the hypothesis that patients with a greater dosage frequency see greater improvement over time is supported by the majority of individual domain models. Specifically, separate LMMs similarly revealed a significant time × dosage frequency group interaction for 9 of the 13 total domains. These 9 domains included the arithmetic, auditory comprehension, auditory memory, naming, quantitative, reading, visual memory, visuospatial, and writing domains (Figure 3; Multimedia Appendix 1). For most of these domains, model results revealed a trend similar to the overall model results, in which a significantly greater rate of change in domain score was observed for higher versus lower practice frequencies (Multimedia Appendix 1). For the arithmetic, auditory comprehension, and auditory memory domains, there was a significantly greater rate of change in domain score for practice frequencies of 2, 3, 4, and ≥5 days per week than that of 1 day

per week (Figure 3A). For the quantitative, reading, visual memory, and visuospatial domains, there was a significantly greater rate of change in the domain score for practice frequencies of 4 and ≥5 days per week than that of 1 day per week (Figure 3B). In the naming and writing domains, despite an overall significant interaction between time and dosage frequency group, the slope was statistically significant only for 4 days per week and 3 days per week (compared with 1 day per week) dosage frequency groups (Multimedia Appendix 1). For the remaining domains—analytical, attention, phonological processing, and production—no significant interaction between time and dosage frequency group was observed, indicating that improvement over the treatment period did not differ based on practice frequency (Figure 4; Multimedia Appendix 1).

Figure 3. Weekly change in domain score as a function of dosage frequency group, by skill domain (significant time × dosage group effect). (A) Arithmetic, auditory comprehension, and auditory memory domains showed a significantly greater rate of change in domain scores for practice frequencies of 2, 3, 4, and ≥5 days per week compared with 1 day per week. (B) Quantitative, reading, visual memory, and visuospatial domains showed a significantly greater rate of change in domain scores for practice frequencies of 4 and ≥5 days per week. compared with 1 day per week.

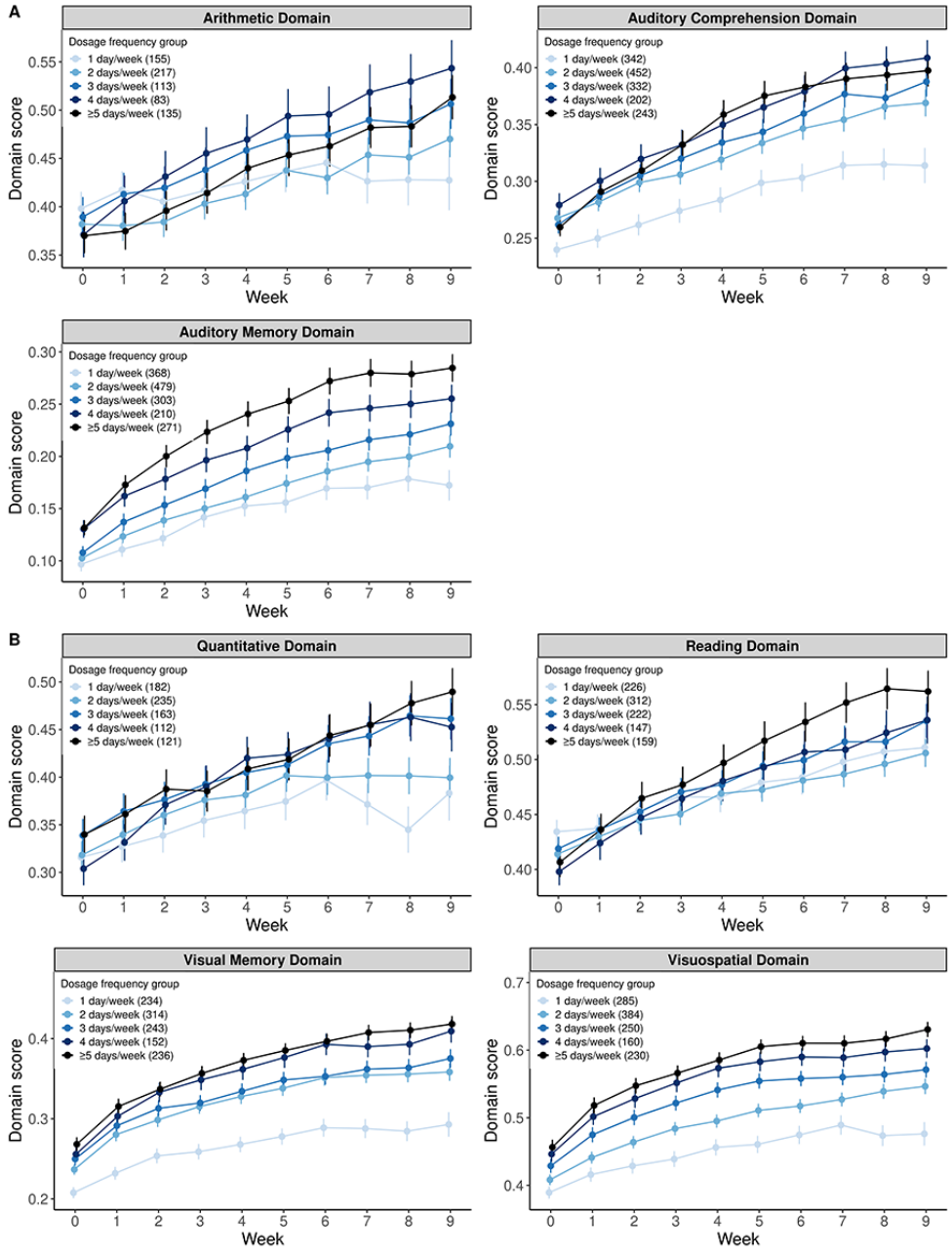
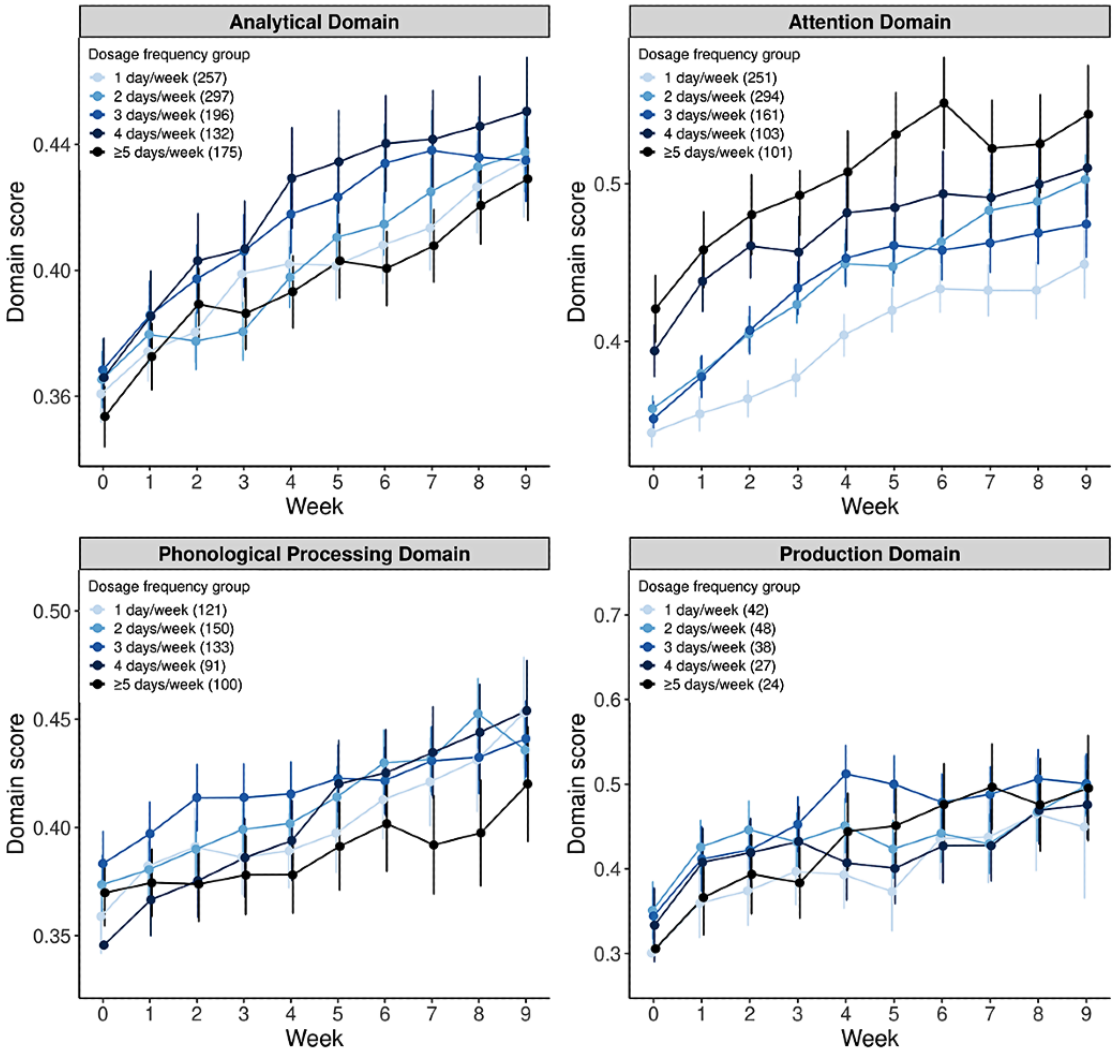


Figure 4. Weekly change in domain score as a function of dosage frequency group, by skill domain (nonsignificant time×dosage group effect).



Discussion

Summary of Findings

Patients with poststroke speech, language, and cognitive impairment generally saw an improvement in their ability to accurately perform tasks of increasing difficulty during their first 10 weeks of self-managed digital therapy across a variety of skill domains. Overall, for most of the individual domains assessed, the rate of improvement was modulated by practice frequency, with significantly greater improvement over time for patients who practiced 3 to 5 days per week than patients who only completed sessions 1 day per week.

These results are generally concordant with the growing body of literature showing the benefits of high-dose and high-intensity SLT on communication outcomes [6,50]. Our work adds to this research by providing results for a large sample over a broad range of functional skill domains. It also underscores the importance of delineating the component parameters of dosage. Increasingly, in the literature, cumulative intervention intensity (ie, total dose) is quantified as the product of session dose, frequency, and intervention duration [29,37,38,51]. In this study, our primary research questions were focused on dosage frequency; but we also accounted for dose amount and duration. Importantly, the results for dosage frequency presented here are independent of overall duration and total number of hours of therapy and therefore underscore the importance of considering practice frequency in addition to other related dose parameters when devising optimal dosage recommendations. Furthermore, the model results showed that although the main effect of the total amount of therapy was significant, it did not predict the rate of improvement over time in the same way that dosage frequency did. This finding demonstrates that, although different dose parameters may be related, they do not necessarily have an equal impact on treatment outcomes.

In this study, analyses were conducted on real-world patient usage data that included a wide range of dosage frequencies, from 1 day per week up to ≥5 days per week. This ensured that the dosage frequencies being investigated are practically achievable. Recent work has identified a major gap between the dose parameters being studied in research—which tend to be uniformly high—and the modest therapy doses being delivered as part of routine clinical practice [51,52]. For example, a study of dosage amounts in a US-based outpatient setting reported a median total therapy dosage of just 7.5 hours over a median 7.7-week treatment duration for individuals with poststroke aphasia, compared with a significantly more intensive dosage regimen (median 20 hours over a median 6-week period)

reported in the aphasia treatment literature from 2009 to 2019 [51]. This dosage gap is a major barrier to the successful clinical implementation of research findings. Thus, investigating naturally occurring dosage frequencies maximizes ecological validity, and by extension, the potential for findings to directly inform clinical dose recommendations. Analysis of a range of dosage frequencies is also important because it allowed for post hoc comparison of individual dosage frequency groups (1 day vs 2 days per week, 2 days vs 3 days per week, etc). The results demonstrated that, across all domains, each additional day of practice per week was associated with a significantly greater improvement over time, with the exception of 4 versus ≥5 days per week. The nonsignificant difference in performance outcome at the upper end of the practice frequency range raises the possibility of diminishing returns, a finding that has also been suggested in other recent work and may be explained by a ceiling effect for certain impairment-based therapies [30]. The existence of a lower threshold for improvement is similarly a source of debate in the limited available literature; for instance, a prior study found no significant differences in outcome for therapy delivered for 48 versus 24 total hours [33]. In contrast, the findings from this study demonstrated significant incremental improvement over time for each additional day of practice, even at the lower end of the practice frequency range. For example, practicing for even 2 days versus 1 day a week confers a modest benefit in treatment outcome, which is useful information for clinicians seeking to set practical and attainable goals for patients.

Taken together, the results of this study provide critical information regarding the optimal dosage for a self-managed digital intervention. Currently, there are few empirical guidelines available to rehabilitation professionals to guide dose prescriptions for any speech-language–focused behavioral intervention, and none of these are specific to self-managed therapy modalities. This study adds to the limited body of existing literature on dose articulation and is the first to focus specifically on dose comparisons for self-managed digital therapy. We anticipate that the results will inform future recommendations of optimal dosage, which is critically needed as the field of speech-language pathology and stroke rehabilitation makes increasing use of digital therapy technologies.

Limitations

This study is not without limitations. For instance, users were not randomly assigned to their dosage frequency group but were binned according to their usage pattern documented in the Constant Therapy system. It is possible that users with less severe impairments (determined by the baseline skill domain score) self-selected into different dosage groups. However, post hoc tests revealed differences in baseline severity between the ≥5 days per week practice group and other practice groups, and not among the 1, 2, 3, or 4 day per week practice groups. Therefore, a difference in baseline severity is not likely to explain the stepwise, incremental effects of dosage frequency found in this study. To further interrogate this question of whether users with less severe impairments are self-selectively getting more exposure to the treatment, we conducted a follow-up correlational analysis comparing the baseline domain score and the total number of practice hours (per skill domain). This analysis revealed no significant relationship between baseline severity and total amount of exposure (Multimedia Appendix 1), indicating that patients with less severe impairments at baseline did not have more exposure to the treatment. Despite the fact that this linear relationship was nonsignificant, all statistical models included both the baseline domain score and total number of hours as covariates to account for any potential effects of baseline severity or total therapy exposure, respectively, on performance gains over the treatment period.

A second limitation of this study was the lack of detailed person-level factors that could influence intervention outcomes. Although the Constant Therapy digital health platform allows for the collection of a large amount of real-world data across several English-speaking countries, it is currently impossible to collect detailed demographic and assessment information from all individuals. Thus, although we have included basic demographic covariates such as age, time since stroke, sex, and a proxy measure for baseline severity in our analysis models, the models would likely be improved with more detailed information about diagnosis, performance on standardized assessment metrics of language (eg, Western Aphasia Battery-Revised) or global function (eg, National Institutes of Health Stroke Scale and Modified Rankin Scale), concurrent medical and cognitive comorbidities, and psychosocial factors. A related limitation is the lack of information available in this data set regarding users' access to direct therapy services. It is likely that for some users, the app-based regimen was used in conjunction with more traditional, in-person SLT, whereas for others, the app constituted the primary or singular mode of therapy. Systematic differences across the dosage groups in amounts of outside (ie, non–app-based) therapy received have the potential to influence observed results, as it is possible that frequent users of the app may also be receiving greater amounts of outside therapy, thus complicating the attribution of performance gain to a greater frequency of in-app practice.

Finally, although designed to be conservative estimates of therapeutic progress, we note that the skill domain scores used in this study are first-order approximations of functioning within a target skill domain. Improved approximations and validation against standardized assessments are the focus of ongoing and future work.

Authors' Contributions

CC, JG, MM, and SK conceptualized the study or contributed to the methodological design. MM and JG contributed the original source code used for database querying and filtering. CC performed the data analysis. CC and SK verified the underlying data

used in this study. All authors contributed to data interpretation. CC and MM wrote the original draft of the manuscript, and all the authors contributed to reviewing and revising the subsequent versions of the manuscript.

Conflicts of Interest

MM, JG, VA, and MA currently receive or have received salary from Constant Therapy Health, producer of the Constant Therapy application. SK currently serves as a scientific adviser and has ownership stock in Constant Therapy Health. CC and SK receive salary from Boston University. Data for this study were analyzed at Boston University as part of a data use agreement with Constant Therapy Health.

Multimedia Appendix 1
The linear mixed-effects models selection process, detailed statistical results, and supplemental analyses.
[DOCX File , 552 KB - jmir_v24i7e36135_app1.docx]

References

1. Aphasia Fact Sheet. National Aphasia Association (NAA). URL: https://www.aphasia.org/aphasia-resources/aphasia-factsheet/ [accessed 2020-12-17]
2. Laska AC, Hellblom A, Murray V, Kahan T, Von Arbin M. Aphasia in acute stroke and relation to outcome. J Intern Med 2001 May;249(5):413-422 [FREE Full text] [doi: 10.1046/j.1365-2796.2001.00812.x] [Medline: 11350565]
3. Paolucci S, Antonucci G, Pratesi L, Traballesi M, Lubich S, Grasso MG. Functional outcome in stroke inpatient rehabilitation: predicting no, low and high response patients. Cerebrovasc Dis 1998;8(4):228-234. [doi: 10.1159/000015856] [Medline: 9684063]
4. Lam JM, Wodchis WP. The relationship of 60 disease diagnoses and 15 conditions to preference-based health-related quality of life in Ontario hospital-based long-term care residents. Med Care 2010 Apr;48(4):380-387. [doi: 10.1097/MLR.0b013e3181ca2647] [Medline: 20220536]
5. Bullier B, Cassoudesalle H, Villain M, Cogné M, Mollo C, De Gabory I, et al. New factors that affect quality of life in patients with aphasia. Ann Phys Rehabil Med 2020 Jan;63(1):33-37 [FREE Full text] [doi: 10.1016/j.rehab.2019.06.015] [Medline: 31352062]
6. Brady MC, Kelly H, Godwin J, Enderby P, Campbell P. Speech and language therapy for aphasia following stroke. Cochrane Database Syst Rev 2016 Jun 01;2016(6):CD000425 [FREE Full text] [doi: 10.1002/14651858.CD000425.pub4] [Medline: 27245310]
7. Code C, Petheram B. Delivering for aphasia. Int J Speech Lang Pathol 2011 Feb;13(1):3-10. [doi: 10.3109/17549507.2010.520090] [Medline: 21329405]
8. Palmer R, Witts H, Chater T. What speech and language therapy do community dwelling stroke survivors with aphasia receive in the UK? PLoS One 2018 Jul 10;13(7):e0200096 [FREE Full text] [doi: 10.1371/journal.pone.0200096] [Medline: 29990345]
9. Bhogal SK, Teasell R, Speechley M. Intensity of aphasia therapy, impact on recovery. Stroke 2003 Apr;34(4):987-993. [doi: 10.1161/01.STR.0000062343.64383.D0] [Medline: 12649521]
10. Worrall LE, Howe T, O'Callaghan A, Hill AJ, Rose M, Wallace SJ, et al. The World Report on Disability as a blueprint for international, national, and local aphasia services. Int J Speech Lang Pathol 2013 Feb;15(1):106-112. [doi: 10.3109/17549507.2012.721004] [Medline: 23101515]
11. Des Roches CA, Kiran S. Technology-based rehabilitation to improve communication after acquired brain injury. Front Neurosci 2017 Jul 28;11:382 [FREE Full text] [doi: 10.3389/fnins.2017.00382] [Medline: 28804443]
12. Doogan C, Dignam J, Copland D, Leff A. Aphasia recovery: when, how and who to treat? Curr Neurol Neurosci Rep 2018 Oct 15;18(12):90 [FREE Full text] [doi: 10.1007/s11910-018-0891-x] [Medline: 30324233]
13. Weidner K, Lowman J. Telepractice for adult speech-language pathology services: a systematic review. Perspect ASHA SIGs 2020 Feb 21;5(1):326-338. [doi: 10.1044/2019_persp-19-00146]
14. Cherney LR, Halper AS, Holland AL, Cole R. Computerized script training for aphasia: preliminary results. Am J Speech Lang Pathol 2008 Feb;17(1):19-34 [FREE Full text] [doi: 10.1044/1058-0360(2008/003)] [Medline: 18230811]
15. Lee JB, Kaye RC, Cherney LR. Conversational script performance in adults with non‐fluent aphasia: treatment intensity and aphasia severity. Aphasiology 2009 Jul;23(7-8):885-897. [doi: 10.1080/02687030802669534]
16. Furnas DW, Edmonds LA. The effect of computerised Verb Network Strengthening Treatment on lexical retrieval in aphasia. Aphasiology 2014 Jan 03;28(4):401-420. [doi: 10.1080/02687038.2013.869304]
17. Fleming V, Brownsett S, Krason A, Maegli MA, Coley-Fisher H, Ong YH, et al. Efficacy of spoken word comprehension therapy in patients with chronic aphasia: a cross-over randomised controlled trial with structural imaging. J Neurol Neurosurg Psychiatry 2020 Nov 05;92(4):418-424 [FREE Full text] [doi: 10.1136/jnnp-2020-324256] [Medline: 33154182]
18. Woodhead ZV, Kerry SJ, Aguilar OM, Ong YH, Hogan JS, Pappa K, et al. Randomized trial of iReadMore word reading training and brain stimulation in central alexia. Brain 2018 Jul 01;141(7):2127-2141 [FREE Full text] [doi: 10.1093/brain/awy138] [Medline: 29912350]

19. De Luca R, Aragona B, Leonardi S, Torrisi M, Galletti B, Galletti F, et al. Computerized training in poststroke aphasia: what about the long-term effects? A randomized clinical trial. J Stroke Cerebrovasc Dis 2018 Aug;27(8):2271-2276. [doi: 10.1016/j.jstrokecerebrovasdis.2018.04.019] [Medline: 29880209]
20. Gerber SM, Schütz N, Uslu AS, Schmidt N, Röthlisberger C, Wyss P, et al. Therapist-guided tablet-based telerehabilitation for patients with aphasia: proof-of-concept and usability study. JMIR Rehabil Assist Technol 2019 Apr 26;6(1):e13163 [FREE Full text] [doi: 10.2196/13163] [Medline: 31025946]
21. Cherney LR, Braun EJ, Lee JB, Kocherginsky M, Van Vuuren S. Optimising recovery in aphasia: learning following exposure to a single dose of computer-based script training. Int J Speech Lang Pathol 2019 Oct;21(5):448-458. [doi: 10.1080/17549507.2019.1661518] [Medline: 31533483]
22. Des Roches CA, Balachandran I, Ascenso EM, Tripodis Y, Kiran S. Effectiveness of an impairment-based individualized rehabilitation program using an iPad-based software platform. Front Hum Neurosci 2014 Jan 5;8:1015 [FREE Full text] [doi: 10.3389/fnhum.2014.01015] [Medline: 25601831]
23. Palmer R, Dimairo M, Cooper C, Enderby P, Brady M, Bowen A, et al. Self-managed, computerised speech and language therapy for patients with chronic aphasia post-stroke compared with usual care or attention control (Big CACTUS): a multicentre, single-blinded, randomised controlled trial. Lancet Neurol 2019 Sep;18(9):821-833 [FREE Full text] [doi: 10.1016/S1474-4422(19)30192-9] [Medline: 31397288]
24. Kurland J, Liu A, Stokes P. Effects of a tablet-based home practice program with telepractice on treatment outcomes in chronic aphasia. J Speech Lang Hear Res 2018 May 17;61(5):1140-1156 [FREE Full text] [doi: 10.1044/2018_JSLHR-L-17-0277] [Medline: 29710115]
25. Godlove J, Anantha V, Advani M, Des Roches C, Kiran S. Comparison of therapy practice at home and in the clinic: a retrospective analysis of the constant therapy platform data set. Front Neurol 2019 Feb 25;10:140 [FREE Full text] [doi: 10.3389/fneur.2019.00140] [Medline: 30858819]
26. Mallet KH, Shamloul RM, Corbett D, Finestone HM, Hatcher S, Lumsden J, et al. RecoverNow: feasibility of a mobile tablet-based rehabilitation intervention to treat post-stroke communication deficits in the acute care setting. PLoS One 2016 Dec 21;11(12):e0167950 [FREE Full text] [doi: 10.1371/journal.pone.0167950] [Medline: 28002479]
27. Munsell M, De Oliveira E, Saxena S, Godlove J, Kiran S. Closing the digital divide in speech, language, and cognitive therapy: cohort study of the factors associated with technology usage for rehabilitation. J Med Internet Res 2020 Feb 07;22(2):e16286 [FREE Full text] [doi: 10.2196/16286] [Medline: 32044752]
28. Harvey SR, Carragher M, Dickey MW, Pierce JE, Rose ML. Treatment dose in post-stroke aphasia: a systematic scoping review. Neuropsychol Rehabil 2021 Dec;31(10):1629-1660. [doi: 10.1080/09602011.2020.1786412] [Medline: 32631143]
29. Warren SF, Fey ME, Yoder PJ. Differential treatment intensity research: a missing link to creating optimally effective communication interventions. Ment Retard Dev Disabil Res Rev 2007;13(1):70-77. [doi: 10.1002/mrdd.20139] [Medline: 17326112]
30. Harvey S, Carragher M, Dickey MW, Pierce JE, Rose ML. Dose effects in behavioural treatment of post-stroke aphasia: a systematic review and meta-analysis. Disabil Rehabil (forthcoming) 2020 Nov 09:1-12. [doi: 10.1080/09638288.2020.1843079] [Medline: 33164590]
31. Bakheit AM, Shaw S, Barrett L, Wood J, Carrington S, Griffiths S, et al. A prospective, randomized, parallel group, controlled study of the effect of intensity of speech and language therapy on early recovery from poststroke aphasia. Clin Rehabil 2007 Oct;21(10):885-894. [doi: 10.1177/0269215507078486] [Medline: 17981847]
32. Breitenstein C, Grewe T, Flöel A, Ziegler W, Springer L, Martus P, FCET2EC study group. Intensive speech and language therapy in patients with chronic aphasia after stroke: a randomised, open-label, blinded-endpoint, controlled trial in a health-care setting. Lancet 2017 Apr 15;389(10078):1528-1538. [doi: 10.1016/S0140-6736(17)30067-3] [Medline: 28256356]
33. Stahl B, Mohr B, Büscher V, Dreyer FR, Lucchese G, Pulvermüller F. Efficacy of intensive aphasia therapy in patients with chronic stroke: a randomised controlled trial. J Neurol Neurosurg Psychiatry 2018 Jun;89(6):586-592 [FREE Full text] [doi: 10.1136/jnnp-2017-315962] [Medline: 29273692]
34. Mozeiko J, Myers EB, Coelho CA. Treatment response to a double administration of constraint-induced language therapy in chronic aphasia. J Speech Lang Hear Res 2018 Jul 13;61(7):1664-1690 [FREE Full text] [doi: 10.1044/2018_JSLHR-L-16-0102] [Medline: 29872835]
35. Harnish SM, Morgan J, Lundine JP, Bauer A, Singletary F, Benjamin ML, et al. Dosing of a cued picture-naming treatment for anomia. Am J Speech Lang Pathol 2014 May;23(2):S285-S299. [doi: 10.1044/2014_AJSLP-13-0081] [Medline: 24686830]
36. Off CA, Griffin JR, Spencer KA, Rogers M. The impact of dose on naming accuracy with persons with aphasia. Aphasiology 2016;30(9):983-1011 [FREE Full text] [doi: 10.1080/02687038.2015.1100705] [Medline: 28133407]
37. Cherney LR. Aphasia treatment: intensity, dose parameters, and script training. Int J Speech Lang Pathol 2012 Oct;14(5):424-431 [FREE Full text] [doi: 10.3109/17549507.2012.686629] [Medline: 22731660]
38. Baker E. Optimal intervention intensity. Int J Speech Lang Pathol 2012 Oct;14(5):401-409. [doi: 10.3109/17549507.2012.700323] [Medline: 22916999]
39. Kiran S, Thompson CK. Neuroplasticity of language networks in aphasia: advances, updates, and future challenges. Front Neurol 2019 Apr 2;10:295 [FREE Full text] [doi: 10.3389/fneur.2019.00295] [Medline: 31001187]

40. Constant Therapy Health. URL: https://constanttherapyhealth.com/constant-therapy/ [accessed 2022-06-22]
41. Braley M, Pierce JS, Saxena S, De Oliveira E, Taraboanta L, Anantha V, et al. A virtual, randomized, control trial of a digital therapeutic for speech, language, and cognitive intervention in post-stroke persons with aphasia. Front Neurol 2021 Feb 12;12:626780 [FREE Full text] [doi: 10.3389/fneur.2021.626780] [Medline: 33643204]
42. Des Roches CA, Mitko A, Kiran S. Relationship between self-administered cues and rehabilitation outcomes in individuals with aphasia: understanding individual responsiveness to a technology-based rehabilitation program. Front Hum Neurosci 2017 Feb 1;11:07 [FREE Full text] [doi: 10.3389/fnhum.2017.00007] [Medline: 28203150]
43. Kiran S, Gerst K, Dubas E. Understanding optimal dosage frequency and patient engagement on improving outcomes using digital therapy. In: International Stroke Conference 2019 Moderated Poster Abstracts. 2019 Presented at: ISC '19; February 6-8, 2019; Honolulu, HI, USA URL: https://www.ahajournals.org/doi/10.1161/str.50.suppl_1.TMP48 [doi: 10.1161/str.50.suppl_1.tmp48]
44. Meteyard L, Davies RA. Best practice guidance for linear mixed-effects models in psychological science. J Mem Lang 2020 Jun;112:104092. [doi: 10.1016/j.jml.2020.104092]
45. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2021. URL: https://www.R-project.org/ [accessed 2021-12-27]
46. 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]
47. Kuznetsova A, Brockhoff PB, Christensen RH. lmerTest package: tests in linear mixed effects models. J Stat Soft 2017;82(13):1-26. [doi: 10.18637/jss.v082.i13]
48. Lenth RV. emmeans: Estimated Marginal Means, aka Least-Squares Means. The Comprehensive R Archive Network. 2021. URL: https://CRAN.R-project.org/package=emmeans [accessed 2021-12-27]
49. Lüdecke D. sjPlot: Data Visualization for Statistics in Social Science. The Comprehensive R Archive Network. 2021 Nov 26. URL: https://CRAN.R-project.org/package=sjPlot [accessed 2021-12-27]
50. REhabilitation and recovery of peopLE with Aphasia after StrokE (RELEASE) Collaborators. Dosage, intensity, and frequency of language therapy for aphasia: a systematic review-based, individual participant data network meta-analysis. Stroke 2022 Mar;53(3):956-967 [FREE Full text] [doi: 10.1161/STROKEAHA.121.035216] [Medline: 34847708]
51. Cavanaugh R, Kravetz C, Jarold L, Quique Y, Turner R, Evans WS. Is there a research-practice dosage gap in aphasia rehabilitation? Am J Speech Lang Pathol 2021 Sep 23;30(5):2115-2129 [FREE Full text] [doi: 10.1044/2021_AJSLP-20-00257] [Medline: 34411485]
52. Skolarus LE, Feng C, Burke JF. No racial difference in rehabilitation therapy across all post-acute care settings in the year following a stroke. Stroke 2017 Dec;48(12):3329-3335 [FREE Full text] [doi: 10.1161/STROKEAHA.117.017290] [Medline: 29089456]

Abbreviations

LMM: linear mixed-effects model
SLT: speech-language therapy

Edited by R Kukafka; submitted 04.01.22; peer-reviewed by E Donoso Brown, J Heo; comments to author 27.04.22; revised version received 18.05.22; accepted 30.05.22; published 20.07.22.

Please cite as:
Cordella C, Munsell M, Godlove J, Anantha V, Advani M, Kiran S
Dosage Frequency Effects on Treatment Outcomes Following Self-managed Digital Therapy: Retrospective Cohort Study
J Med Internet Res 2022;24(7):e36135
URL: https://www.jmir.org/2022/7/e36135
doi:10.2196/36135
PMID:35857353

Original Paper

Providing Brief Personalized Therapies for Insomnia Among Workers Using a Sleep Prompt App: Randomized Controlled Trial

Tomonari Shimamoto[1], DPH; Ryuji Furihata[2], PhD, MD; Yukako Nakagami[2], PhD, MD; Yukiko Tateyama[1], PhD; Daisuke Kobayashi[2], PhD, MD; Kosuke Kiyohara[3], DPH; Taku Iwami[1], PhD, MD

[1]Department of Preventive Services, School of Public Health, Graduate School of Medicine, Kyoto University, Kyoto, Japan

[2]Agency for Health, Safety and Environment, Kyoto University, Kyoto, Japan

[3]Department of Food Science, Faculty of Home Economics, Otsuma Women's University, Tokyo, Japan

Corresponding Author:
Ryuji Furihata, PhD, MD
Agency for Health, Safety and Environment
Kyoto University
Yoshida honmachi, Sakyo-ku
Kyoto, 606-8501
Japan
Phone: 81 75 753 2428 ext 2428
Email: furihata.ryuji.2x@kyoto-u.ac.jp

Abstract

Background: Insomnia is the most common sleep disorder and the foremost health concern among workers. We developed a new sleep prompt app (SPA) for smartphones to positively alter the users' consciousness and behavior by sending timely short messages for mild sleep problems at an early stage.

Objective: The aim of this study is to investigate the effectiveness of the SPA in providing brief personalized therapy for insomnia among workers.

Methods: We conducted a 2-arm parallel randomized controlled trial. The intervention group used the SPA, and the control group received no intervention. Participants were recruited between November 2020 and January 2021. The researcher sent emails for recruitment to more than 3000 workers of 2 companies and 1 university in Japan. The SPA provided personalized prompt messages, sleep diaries, sleep hygiene education, stimulus control therapy, and sleep restriction therapy. The prompt messages were sent automatically to the participants to encourage them to improve their sleep habits and sleep status and were optimized to the individual's daily rhythm. The intervention program duration was 4 weeks. The primary outcome was a change in the Insomnia Severity Index (ISI) for the study period. The ISI was obtained weekly using a web questionnaire.

Results: A total of 116 Japanese workers (intervention group n=60, control group n=56) with sleep disorders were recruited. Two participants in the intervention group were excluded from the analyses because of challenges in installing the SPA. The mean ISI scores at baseline were 9.2 for both groups; however, after 4 weeks, the mean ISI scores declined to 6.8 and 8.0 for the intervention and control groups, respectively. Primary analysis using a linear mixed model showed a significant improvement in the temporal trends of the ISI in the SPA group and in the total population (P=.03). Subgroup analyses of ISI-8-insomniacs revealed a significant improvement in the temporal trends of ISI in the SPA group (P=.01), and the CFS score for physical condition significantly improved following the intervention (P=.02).

Conclusions: This study demonstrates the effectiveness of the SPA in providing brief personalized therapy for insomnia among Japanese workers with mild insomnia. The physical fatigue score significantly improved in ISI-8-insomniacs. Thus, SPA could play an important role in reducing the adverse effects of sleep disorders in workers. To promote the wide use of the SPA in the future, further studies are required to examine its effectiveness in other age groups and individuals with health problems.

Trial Registration: University Medical Information Network Clinical Trials Registry (UMIN-CTR) UMIN000042263; https://center6.umin.ac.jp/cgi-open-bin/ctr_e/ctr_view.cgi?recptno=R000046295

(J Med Internet Res 2022;24(7):e36862) doi:10.2196/36862

KEYWORDS

sleep prompt app; smartphone; brief personalized therapies for insomnia; worker; randomized controlled trial; Japan

Introduction

Insomnia is the most common sleep disorder, and previous epidemiological studies revealed that almost 20% of the general adult population exhibits symptoms of insomnia [1,2]. Insomnia is associated with increased risks of mental disorders such as depression, anxiety, alcohol abuse, and psychosis [3] and physical diseases including hypertension, hyperglycemia, hyperlipidemia, obesity, and cardio-cerebrovascular events [4,5]. In workers, insomnia is considered an important problem because it is associated with reduced work performance, processing errors, accidents at work, absenteeism, reduced quality of life, and symptoms of depression even if the insomnia is mild [6]. Thus, insomnia has become one of the foremost health concerns among workers.

In the clinical field, cognitive behavioral therapy for insomnia (CBT-I) or brief therapy for insomnia (BTI), which combine multiple treatment techniques, has been shown to be effective in randomized control trials (RCTs), and these therapies are highly recommended (strong for CBT-I, conditional for BTI) for the treatment of insomnia in the US guidelines [7]. Moreover, the development of CBT-I or BTI using internet and smartphone applications has progressed, and they have been found to be as effective as treatments provided by trained clinicians [8-10]. Therefore, CBT-I or BTI using internet and smartphone applications is expected to be widely used in the health care field through further technological innovation. However, due to the use of standardized educational programs, it has been difficult to provide individualized programs, and sometimes other media or telephone support has been needed to address individual problems [11]. Furthermore, the effectiveness of the program in the health care field for mild cases has not been fully examined.

We developed a new technology to change the users' consciousness and behavior in a desirable direction by sending short messages named “prompt messages” that encourage behavioral change in line with the users' daily rhythms in using mobile devices (eg, smartphones) and by promoting awareness of lifestyle improvement. If we can detect individual sleep problems (insufficient sleep time, excessive daytime sleepiness, disturbance of sleep-wake rhythms, issues requiring improvement in sleep hygiene, a discrepancy between total time in bed and total sleep time, and reduced sleep efficiency) and use prompting technology to provide notifications that lead to sleep improvement in real-time according to an individual's daily rhythm as if a sleep specialist were present when needed, we can provide highly individualized and efficient interventions for improved sleep behavior. Hence, to provide real-time interventions, we developed a new sleep prompt app (SPA) for smartphones through industry-academia collaboration. The SPA is expected to enable early detection and intervention in mild sleep disorders.

The primary aim of this study is to investigate whether SPA can aid in improving insomnia among Japanese workers by providing brief personalized therapy.

Methods

Recruitment

The study was designed as a 2-arm (intervention using the SPA vs wait-list control) parallel RCT. Participants were recruited between November 2020 and January 2021. All follow-up was completed by March 2021. An email for recruitment was sent to more than 3000 workers of 2 companies and 1 university in Japan. The institutional affiliations (Kyoto University; HealthTech Laboratory, Inc; and OKI Electric Industry Co, Ltd) were displayed in the emails as the organizations conducting the study. To avoid the risk of divulging information to competing research institutions, we selected companies and universities affiliated to our research group for recruitment of participants. However, we excluded workers of departments involved in the study. Both eligible employees and their acquaintances were allowed to participate in the study. Written informed consent was obtained through participation in face-to-face information sessions or online information sessions using a web conference system. After the information sessions, all subsequent communications were conducted via email.

The inclusion criteria were workers aged 20 years or older, ownership of a smartphone, and an affirmative answer to the question “Do you have any sleep problems?” that was asked during the application for participation in this study.

There were 6 exclusion criteria: (1) individuals whose version or model of smartphone did not support the SPA, (2) shift workers, (3) individuals likely to be subjected to significant time zone changes such as overseas travel during the study period, (4) those who had a total score of 10 or more on the Patient Health Questionnaire-9 at the time of application for participation in the study, (5) those who responded that the frequency of “thoughts that you would be better off dead or of hurting yourself in some way” was more than half of the days in the previous 2 weeks, and (6) those who were judged to be inappropriate for the study.

Randomization

Random assignment into the SPA group or nonintervention control group was performed by stratified block allocation with an allocation ratio of 1:1, a block size of 4 using sex and 3 groups of total Insomnia Severity Index (ISI) scores obtained at the time of study entry: <8, 8-14, and ≥15. Central randomization with computer-generated tables was used for the allocation. One investigator (TS) prepared the list, and another investigator (KK) independently performed the allocation. Due to the nature of the intervention, the participants were not blinded to their allocation. The researchers included statisticians (KK), and data analysts (RF and YN) who were blinded to the assignment of the participants to their respective groups.

Interventions

Overview

The SPA consisted of a smartphone app (HealthTech Laboratory, Inc) and a prompt notification server (OKI Electric Industry Co, Ltd). For the intervention, the smartphone app, Kenko-Nikki (Health Diary), which is a personal health

record–related app developed by HealthTech Laboratory, Inc, a Kyoto University–originated venture company in the Kyoto University Incubation Program [12], was customized for this study. Following the manufacturer's instructions, participants in the intervention group installed the app (TestFlight for iOS or DeployGate for Android) to use the SPA for free. The brief personalized therapy for insomnia used in the intervention consisted of 5 components: (1) prompt messages, (2) a sleep diary, (3) sleep hygiene education, (4) stimulus control therapy, and (5) sleep restriction therapy.

Each study lasted 28 days, and we conducted 7 trials between November 2020 and February 2021. The intensity of use or dosage for the SPA was not recorded in the research data. The participants in the control group received the SPA after completing the posttest questionnaire. During the study period, an email contact was provided for inquiries.

Prompt Messages

A sleep medicine specialist (RF) created more than 200 types of prompt messages, the timing of the message sending (after waking, around noon, in the evening, before bedtime), and the prompt messages' sending conditions according to sleep habits and sleep states. Prompt messages and their transmission conditions were programmed and set up on the prompt notification server at OKI Electric Industry Co, Ltd. At the beginning of the study, information on sex, age, ISI, sleep habits, and list of times when each participant would be more (or less) receptive to prompt messages obtained from the online questionnaire was entered into the prompt notification server; data from the sleep diary, the ISI scores from weekly questionnaires, and sleep schedules set by the participant were also entered. Based on these participant data, the prompt notification server selected an appropriate prompt message and time at which participants would be receptive to the information and then automatically sent the prompt message to the participant at the optimal time. The dispatched prompt messages were received by the participants' SPA (Figure S1, Multimedia Appendix 1).

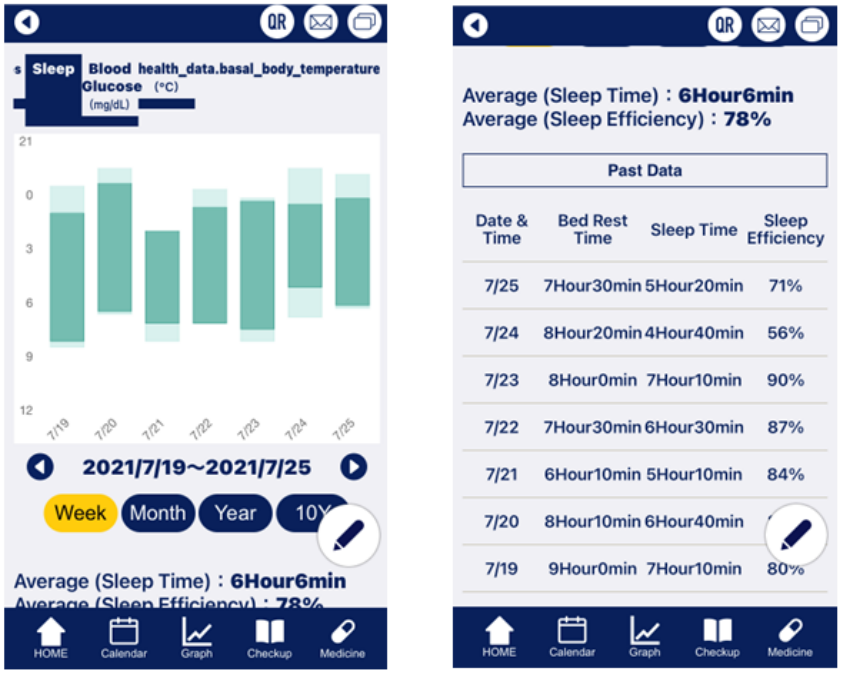
The prompt messages were sent with illustrations of female characters according to their content. The prompt messages consisted of comments related to sleep diaries, sleep hygiene education, stimulus control therapy, or sleep restriction therapy and were sent several times in a day on an everyday basis. An example of a prompt message is shown in Figure 1.

Figure 1. Sleep prompt app prompt message.

Sleep Diary Measures

The participants filled out a sleep diary daily using the SPA. In the sleep diary, they recorded the following 10 items: (1) the time of getting into bed; (2) sleep-onset latency; (3) number of awakenings; (4) duration of awakenings; (5) time of final awakening; (6) bed-out latency; (7) duration of napping; (8) time of day of feeling excessive sleepiness during daytime; (9) behavior before bedtime, including alcohol consumption, caffeine intake, smoking, and hypnotic medication use; and (10) whether the recording date was a weekday or a holiday. From these variables, the SPA calculated the time in bed (time in bed = final arising time – time of going to bed), sleep time (total sleep time = time in bed – sleep-onset latency – wake after sleep onset – bed-out latency), and sleep efficiency (sleep efficiency = [total sleep time/time in bed] × 100). In addition, the SPA presented these sleep variables in numerical and graphical forms (Figure 2). When the participants consistently entered data in their sleep diary, they received prompt messages commending them.

Figure 2. Screenshot of the sleep diary in the sleep prompt app.



Sleep Hygiene Education and Stimulus Control Therapy

Prompt messages were sent for helpful advice on lifestyle changes related to daily rhythm, light exposure, sound, temperature, diet intake, exercise, bathing, drinking, smoking, caffeine consumption, sleep duration, napping, and emotional distress. Prompt messages of sleep hygiene instructions were sent from week 1 to 4. Recommended sleep-related behaviors based on the stimulus control method were delivered as prompt messages at weeks 2 to 4.

Sleep Restriction Therapy

At the end of the week, the participants set their sleep schedules for the next week. First, they entered their desired wake-up times on the SPA. The SPA presented the recommended bedtime, which was calculated by subtracting the previous week's "average total sleep time + 30 minutes" from the desired wake-up time. The participants set their desired bedtime according to the recommended bedtime. The recorded sleep schedules in the sleep diary were entered into the prompt notification server, and based on this information, the prompt notification server sent a prompt message containing advice on actual wake-up time, bedtime, and sleep efficiency. When a participant's sleep efficiency was >90%, the SPA allowed 15 minutes extra time in bed. When a participant's average sleep efficiency was 85%-90%, the SPA suggested the same. When a participant's average sleep efficiency was <85%, the SPA suggested a further restriction of 15 minutes. The sleep restriction method was implemented in weeks 2-4.

Measurements From the Online Questionnaire

All online questionnaires were distributed using Google Forms. The email which included information of the questionnaire and the URL of the questionnaire were sent to all participants.

Primary Measure

We administered the Japanese version of the ISI questionnaire [13,14]. This is a 7-item questionnaire with scores ranging from 0 (no insomnia) to 28 (severe insomnia). Individuals with scores ≥8 were defined as ISI-8-insomniacs [13].

Secondary Measures

Chalder Fatigue Scale

The Chalder fatigue scale (CFS) is a self-administered questionnaire for measuring the extent and severity of fatigue [15]. The items are benign and nonthreatening, asking about sensations and functionality—such as "Do you have problems starting things?" and "Do you have difficulty concentrating?"—rather than about any beliefs or opinions about health status. Each of the 11 items is answered on a 4-point scale ranging from asymptomatic to maximum symptomologies, such as "better than usual," "no worse than usual," "worse than usual," and "much worse than usual." According to Likert scoring, responses on the extreme left received a score of 0, increasing to 1, 2, or 3 as the participants became more symptomatic. The respondents' global scores ranged from 0 to 33. The global score also includes 2 dimensions: physical fatigue (measured by items 1-7) and psychological fatigue (measured by items 8-11).

The World Health Organization Health and Work Performance Questionnaire Short Form

Work performance was assessed using the World Health Organization Health and Work Performance Questionnaire (HPQ) short form [16,17]. The HPQ includes a self-reporting assessment of the on-the-job work performance of most workers in the same job (B9), actual performance within 1 or 2 years before the survey (B10), and actual performance within 4 weeks (28 days) before the survey (B11). The on-the-job work performance scale is a 0-10 self-anchoring scale in which 0

indicates the worst performance at the workplace and 10 represents the performance of a top employee. Although several methods have been created to assess work performance, we evaluated relative presenteeism and absolute presenteeism [17]. We calculated relative presenteeism by a ratio of actual performance (B11) to the performance of most workers at the same job (B9, possible performance). Regarding the distribution of relative presenteeism, 0.25 is the worst relative performance (25% or less of other workers' performance) and 2.0 is the best performance (200% or more of other workers' performance) [17]. We calculated absolute presenteeism as follows: 10 × actual performance within 1 or 2 years before the survey (B11) [17].

Other Measures

The online questionnaire included questions regarding sociodemographic information (sex and age), height, weight, caffeine consumption (calculated per day as 95 mg per cup of coffee, 55 mg per cup of tea, and 45 mg per cup of cola), alcohol consumption (never, sometimes, or every day), smoking habits (yes or no), living with others or alone, hypnotic medication (more than once a week or not), and use of alcohol as an aid to sleep (more than once a week or not). BMI was calculated using height and weight, with 2 groups being created using a cutoff of 25: obese (BMI≥25 kg/m^2) or normal.

Statistical Analysis

Sample-size calculations were based on our pilot study (not published) using a prototype app; the intervention produced a 1.3-point reduction (SD 2.3) on the ISI. We needed 50 participants per intervention group (100 in total) to detect 80% power using a 2-sided α of .05. Accounting for 20% potential attrition, we estimated the need to enroll 120 participants.

Descriptive statistics were computed for participant characteristics. We computed the differences in clinical and demographic variables between the intervention and control groups by original assigned groups using the independent samples *t* test for continuous variables and chi-square tests for categorical variables. As a primary analysis, linear mixed-effects models were applied to longitudinal data to measure the effectiveness of the intervention, to account for repeated measurements over time, and to avoid the imputation of missing data. These models included the ISI scores of the 2 groups, 5 time points (baseline, and 1, 2, 3, and 4 weeks later), and the time × intervention interaction. As secondary measures, we assessed the changes in absolute presenteeism score, relative presenteeism score, CFS total 11, CFS mental 11, and CFS physical 11 with an independent samples *t* test. A subgroup analysis was conducted for ISI-8-insomniacs. No interim analysis was conducted. Statistical significance was defined as $P<.05$ using 2-tailed tests. All analyses were performed using SPSS (version 27.0, IBM Corp) for Windows (Microsoft Corp).

Ethical Considerations

The study protocol was approved by the Ethics Committee of Kyoto University (#C1478). The study was conducted in accordance with the Declaration of Helsinki. After identification at the information session and provision of written informed consent forms under real names, a correspondence chart between IDs and real names was created, and the correspondence chart was kept in strict confidence at the research department of Kyoto University. All data collected from participants' responses to the web questionnaire and the app were managed in a manner that was linked to their IDs, thus guaranteeing participants' privacy.

Results

The CONSORT (Consolidated Standards of Reporting Trials) flow diagram of participant recruitment is shown in Figure 3. Of the 215 participants who participated in the study, 151 were eligible to participate in the trial briefing session, with 63 being excluded for not meeting the inclusion criteria. People with severe complications (late pregnancy and an individual with hypersomnia) and 1 who canceled participation in the trial briefing session were excluded. A further 26 were excluded because they could not attend the trial briefing session, and 7 refused to participate. Hence, 118 were eligible for allocation: 62 were assigned to the intervention group and 56 to the control group. Of the 62 participants assigned to the intervention group, 2 participants who failed to install the app were excluded from the analysis as dropouts. Finally, 116 participants were included in the analysis. No adverse effects were observed in either group. Although the number of participants did not reach 120, recruitment was terminated because the dropout rate was below the estimated number.

The participants' characteristics (total population and ISI-8-insomniacs) are shown in Table 1. In the total population, the percentage of alcohol use as a sleep aid was significantly different between the 2 groups (χ^2=4.22; P=.04). In ISI-8-insomniacs, the percentage of alcohol use (χ^2_2 =6.81; P=.03), current smoking (χ^2_1=3.90; P=.05), and alcohol use as a sleep aid (χ^2_1=5.85; P=.02) was significantly different between the 2 groups.

Figure 4 shows the ISI scores at 4 weeks for the total population. The mean ISI scores for the SPA and control groups were 9.2 and 9.2 at baseline, 7.4 and 7.7 at 1 week, 7.0 and 7.7 at 2 weeks, 6.5 and 7.9 at 3 weeks, and 6.8 and 8.0 at 4 weeks, respectively. Statistical analysis using a linear mixed model showed that the difference in the mean ISI scores of both groups was significantly different (P=.03). The mean difference in the change in ISI scores of the 2 groups at baseline and after 4 weeks was 1.14 points (95% CI –2.31 to 0.04).

Figure 5 shows the ISI scores of the ISI-8-insomniacs at week 4. The mean ISI scores for the SPA and control groups were 11.3 and 11.6 at baseline, 8.6 and 9.5 at 1 week, 8.0 and 9.6 at 2 weeks, 7.4 and 9.9 at 3 weeks, and 7.5 and 9.7 at 4 weeks, respectively. Statistical analysis using a linear mixed model showed that the change in ISI scores for both groups was significantly different (P=.01). The mean difference in the change of ISI scores between the 2 groups at baseline and after 4 weeks was 1.79 points (95% CI –3.31 to –0.26).

The scores of presenteeism and fatigue at baseline and postintervention (total population and ISI-8-insomniacs) are shown in Table 2. In the total population, none of the variables

showed a statistically significant difference. In ISI-8-insomniacs, physical CFS showed significant improvement.

The effect of SPA on sleep parameters in the sleep diary is shown in Multimedia Appendix 2 (Table S1).

Figure 3. The CONSORT (Consolidated Standards of Reporting Trial) flow diagram of recruitment.

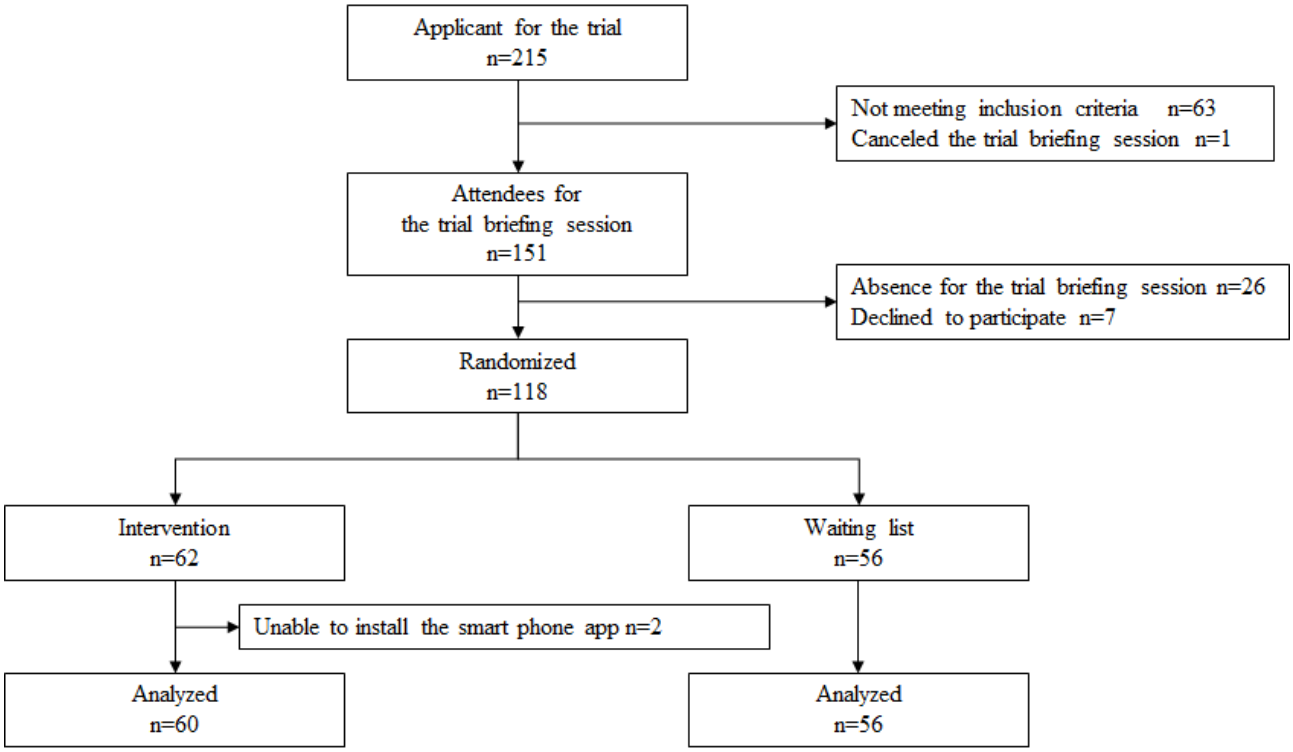


Table 1. Participant characteristics (total population and ISI-8-insomniacs).

Variables	SPA[a]	Control
Total population[b]		
Age (years), mean (SD)	38.6 (11.3)	42.7 (11.5)
Female sex, n (%)	21 (35.0)	17 (30.4)
Live-in alone, n (%)	18 (30.0)	13 (23.2)
Obese (BMI ≥25 kg/m^2), n (%)	12 (20.0)	8 (14.3)
Alcohol use, n (%)		
Never	11 (18.3)	19 (33.9)
Sometimes	35 (58.3)	25 (44.6)
Everyday	14 (23.3)	12 (21.4)
Not current smoker, n (%)	55 (91.7)	54 (96.4)
Caffeine consumption, mean (SD)	191.7 (124.1)	177.3 (138.5)
Do not use alcohol as sleep-aid, n (%)	59 (98.3)	50 (89.3)
Do not use hypnotic medication, n (%)	58 (96.7)	53 (94.6)
ISI[c]-8-insomniacs[d]		
Age (years), mean (SD)	37.7 (11.7)	45.0 (10.7)
Female sex, n (%)	15 (40.5)	11 (32.4)
Live-in alone, n (%)	10 (27.0)	9 (26.5)
Obese (BMI ≥25 kg/m^2), n (%)	7 (18.9)	7 (20.6)
Alcohol use, n (%)		
Never	6 (16.2)	15 (44.1)
Sometimes	23 (62.2)	13 (38.2)
Everyday	8 (21.6)	6 (17.6)
Not current smoker, n (%)	33 (89.2)	34 (100.0
Caffeine consumption, mean (SD)	195.0 (137.9)	177.8 (132.7)
Do not use alcohol as sleep-aid, n (%)	37 (100.0)	29 (85.3)
Do not use hypnotic medication, n (%)	35 (94.6)	31(91.2)

[a]SPA: sleep prompt app.

[b]The SPA and control groups had 60 and 56 participants, respectively.

[c]ISI: Insomnia Severity Index.

[d]The SPA and control groups had 37 and 34 participants, respectively.

Figure 4. Trend in the ISI score during the 4-week trial period in the total population. ISI: Insomnia Severity Index; SPA: sleep prompt app.

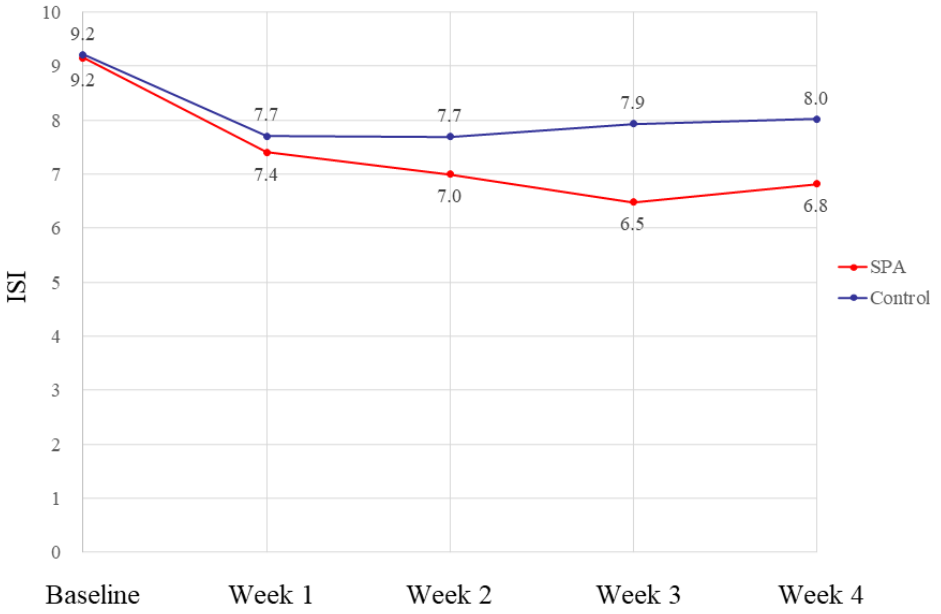


Figure 5. Trend in the ISI score during the 4-week trial period in the population with ISI-8-insomniacs. ISI: Insomnia Severity Index; SPA: sleep prompt app.

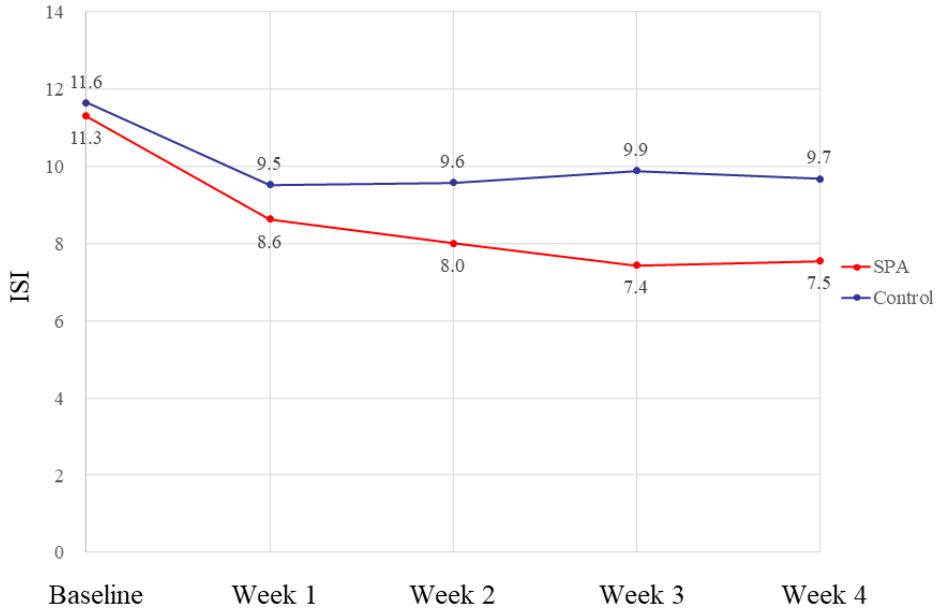


Table 2. Changes of presenteeism scores and fatigue scales before and after the intervention in the total population and ISI-8-insomniacs.

Variables	Baseline, mean (SD)	Posttest, mean (SD)	Comparison with baseline		*P* value
			Mean (SD)	95% CI	
Total population					
CFS[a] total 11					
SPA[b]	15.58 (5.43)	13.35 (5.45)	–2.23 (4.35)	–2.45 to 0.87	.35
Control	16.36 (4.77)	14.91 (5.29)	–1.45 (4.68)	N/A[c]	N/A
CFS mental 11					
SPA	6.08 (1.87)	4.45 (2.27)	–1.63 (1.75)	–0.83 to 0.52	.66
Control	6.50 (1.88)	5.02 (2.04)	–1.48 (1.93)	N/A	N/A
CFS physical 11					
SPA	9.50 (4.11)	8.90 (3.64)	–0.60 (3.21)	–1.89 to 0.62	.32
Control	9.86 (3.49)	9.89 (3.87)	0.04 (3.59)	N/A	N/A
Relative presenteeism score					
SPA	0.95 (0.25)	0.98 (0.21)	0.03 (0.30)	–0.08 to 0.15	.52
Control	0.93 (0.28)	0.93 (0.23)	0.00 (0.30)	N/A	N/A
Absolute presenteeism score					
SPA	64.00 (16.59)	68.33 (15.09)	4.33 (15.44)	–4.66 to 6.90	.70
Control	59.29 (15.36)	62.50 (15.40)	3.21 (15.97)	N/A	N/A
ISI[d]8-insomniacs					
CFS total 11					
SPA	17.65 (4.85)	14.38 (5.58)	–3.27 (3.73)	–4.00 to 0.11	.06
Control	17.12 (4.57)	15.79 (5.76)	–1.32 (4.92)	N/A	N/A
CFS mental 11					
SPA	6.54 (1.77)	4.68 (2.42)	–1.86 (1.77)	–1.16 to 0.72	.64
Control	6.74 (1.91)	5.09 (2.31)	–1.65 (2.19)	N/A	N/A
CFS physical 11					
SPA	11.11 (3.70)	9.70 (3.63)	–1.41 (2.71)	–3.21 to –0.25	.02
Control	10.38 (3.23)	10.71 (4.10)	0.32 (3.51)	N/A	N/A
Relative presenteeism score					
SPA	0.95 (0.26)	0.98 (0.24)	0.03 (0.35)	–0.10 to 0.21	.46
Control	0.90 (0.30)	0.87 (0.23)	–0.03 (0.31)	N/A	N/A
Absolute presenteeism score					
SPA	61.35 (16.19)	67.30 (14.07)	5.95 (12.35)	–3.77 to 10.37	.36
Control	56.76 (15.71)	59.41 (16.69)	2.65 (17.29)	N/A	N/A

[a]CFS: Chalder fatigue scale.

[b]SPA: sleep prompt app.

[c]N/A: not applicable.

[d]ISI: Insomnia Severity Index.

Discussion

Principal Results

In this RCT, we examined the effectiveness of the new SPA in providing brief personalized therapy for relatively mild insomnia among Japanese workers. The results showed that the SPA significantly improved the primary outcome of ISI scores in both the total population and the ISI-8-insomniacs. In addition, the SPA also improved physical fatigue as measured by the physical CFS in ISI-8-insomniacs. To our knowledge, this is the first study to demonstrate the effectiveness of SPA in treating insomnia among workers.

Limitations

There were several methodological limitations to our study. First, we used a subjective rating scale. The use of an objective assessment of sleep, such as polysomnography, would strengthen the results of this study. Although some studies have reported that self-reported data on sleep status concur with physiological data [18,19], using objective data (ie, physiologic measurements such as polysomnography) is desirable in future studies. Second, the participants were not blinded. The participants were aware of their assignment to the intervention or control group, and this difference could have affected the outcome. Because only those in the intervention group used the app and maintained a record in their sleep diaries, the placebo effect might have affected their behavior and influenced the results. Third, the selection of companies and universities from which participants were recruited was not random, and therefore sampling bias may exist. Fourth, the long-term prognosis of the intervention was not investigated.

Comparison With Prior Work

Previous studies demonstrated that internet-based CBT-I improved ISI scores. One RCT compared the effectiveness of web-based CBT-I with telehealth-based CBT for chronic insomnia and found that both web- and telehealth-based treatment improved the ISI score [20]. Another RCT showed that internet-delivered CBT-I improved the ISI scores for patients with insomnia or depression [9]. In our study, the SPA group showed a statistically significant decrease in ISI score, but the difference in ISI scores between the 2 groups was smaller than that reported in previous studies [10]. In a meta-analysis of 2 previous studies, the group receiving internet-based CBT showed a decrease in the ISI score of 2.28 points (–2.89 to –1.67) compared with the control group [10]. There are some possible explanations for this. First, the low ISI scores of the study participants at baseline could have affected the results. We found a large change in the ISI scores of ISI-8-insomniacs. The fact that we were able to show the efficacy of the SPA for mild insomnia in this study is important for the establishment of future preventive methods through early interventions for sleep disorders among workers. Second, in some of the previous studies, email or phone support was required in addition to the app [9,20]. However, considering the need for widespread implementation, it is important that the SPA is effective independently, without the need of additional supportive measures.

In this study, the dropout rate was 3.2% (n=2), and most participants were able to complete the final questionnaire. In many previous studies, about 20% to 40% of the participants dropped out [11]. The app used in this study sends prompt messages to the participants at a convenient time for them, which might have contributed to maintaining their motivation and reducing the number of dropouts.

In this study, a significant improvement in physical fatigue was observed among the ISI-8-insomniacs. Fatigue symptoms have been reported as one of the most frequent daytime complaints of patients with insomnia [21], which has a major impact on day-to-day functioning and quality of life [22]. Clinical studies on patients with chronic obstructive pulmonary disease and on cancer survivors have shown that CBT-I significantly improves fatigue and insomnia in patients with insomnia [23,24]. Although no previous study has examined the effect of internet-delivered CBT-I or BTI on recovery from fatigue in workers, our intervention with the SPA showed, for the first time, the possibility of recovery from physical fatigue.

Conclusions

This RCT showed the effectiveness of the SPA in providing brief personalized therapy for treating relatively mild insomnia among Japanese workers. The physical fatigue score was also significantly improved in ISI-8-insomniacs. The SPA could play an important role in reducing the adverse effects of sleep problems in communities. To promote the use of SPA in the future, further studies are warranted to examine its effectiveness in other age groups and individuals with health problems. To increase generalizability, future studies with follow-up surveys are required to examine the long-term prognosis of the intervention.

Acknowledgments

We would like to thank the individuals that participated in this study. We would like to thank Editage for English-language editing. This study was supported by a joint research grant from HealthTech Laboratory, Inc; OKI Electric Industry Co, Ltd; and Kyoto University. RF, TS, and TI were supported by the Japan Society for the Promotion of Science (KAKENHI grant #JP 20K07967).

Authors' Contributions

TS, RF, YN, YT, DK, KK, and TI contributed to study design and data interpretation. TS, YT, and DK contributed to data acquisition. RF, YN, and KK were involved in data analysis. All authors critically revised and approved the manuscript to be published.

Conflicts of Interest

TI is an unpaid outside director of HealthTech Laboratory, Inc, and receives grants from the Kyoto University Health Service and HealthTech Laboratory Inc, Joint Research Fund. TS and YT are employed by the Kyoto University Health Service and Health Tech Laboratory, Inc, Joint Research Fund. The other authors have no conflicts of interest to declare.

Multimedia Appendix 1
Overview of the sleep prompt app.
[PNG File , 2842 KB - jmir_v24i7e36862_app1.png]

Multimedia Appendix 2
Effect of the sleep prompt app on the sleep parameters of the sleep diary.
[DOCX File , 17 KB - jmir_v24i7e36862_app2.docx]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1250 KB - jmir_v24i7e36862_app3.pdf]

References

1. Furihata R, Uchiyama M, Takahashi S, Suzuki M, Konno C, Osaki K, et al. The association between sleep problems and perceived health status: a Japanese nationwide general population survey. Sleep Med 2012 Aug;13(7):831-837. [doi: 10.1016/j.sleep.2012.03.011] [Medline: 22609021]
2. Kim K, Uchiyama M, Okawa M, Liu X, Ogihara R. An epidemiological study of insomnia among the Japanese general population. Sleep 2000 Feb 01;23(1):41-47. [Medline: 10678464]
3. Hertenstein E, Feige B, Gmeiner T, Kienzler C, Spiegelhalder K, Johann A, et al. Insomnia as a predictor of mental disorders: a systematic review and meta-analysis. Sleep Med Rev 2019 Feb;43:96-105. [doi: 10.1016/j.smrv.2018.10.006] [Medline: 30537570]
4. He Q, Zhang P, Li G, Dai H, Shi J. The association between insomnia symptoms and risk of cardio-cerebral vascular events: A meta-analysis of prospective cohort studies. Eur J Prev Cardiol 2017 Jul;24(10):1071-1082. [doi: 10.1177/2047487317702043] [Medline: 28359160]
5. Zhang Y, Jiang X, Liu J, Lang Y, Liu Y. The association between insomnia and the risk of metabolic syndrome: A systematic review and meta-analysis. J Clin Neurosci 2021 Jul;89:430-436. [doi: 10.1016/j.jocn.2021.05.039] [Medline: 34074598]
6. Richter K, Acker J, Adam S, Niklewski G. Prevention of fatigue and insomnia in shift workers-a review of non-pharmacological measures. EPMA J 2016;7:16 [FREE Full text] [doi: 10.1186/s13167-016-0064-4] [Medline: 27486484]
7. Edinger JD, Arnedt JT, Bertisch SM, Carney CE, Harrington JJ, Lichstein KL, et al. Behavioral and psychological treatments for chronic insomnia disorder in adults: an American Academy of Sleep Medicine clinical practice guideline. J Clin Sleep Med 2021 Feb 01;17(2):255-262 [FREE Full text] [doi: 10.5664/jcsm.8986] [Medline: 33164742]
8. Espie CA, Kyle SD, Williams C, Ong JC, Douglas NJ, Hames P, et al. A randomized, placebo-controlled trial of online cognitive behavioral therapy for chronic insomnia disorder delivered via an automated media-rich web application. Sleep 2012 Jun;35(6):769-781 [FREE Full text] [doi: 10.5665/sleep.1872] [Medline: 22654196]
9. Blom K, Jernelöv S, Kraepelien M, Bergdahl MO, Jungmarker K, Ankartjärn L, et al. Internet treatment addressing either insomnia or depression, for patients with both diagnoses: a randomized trial. Sleep 2015 Feb 01;38(2):267-277 [FREE Full text] [doi: 10.5665/sleep.4412] [Medline: 25337948]
10. Seyffert M, Lagisetty P, Landgraf J, Chopra V, Pfeiffer PN, Conte ML, et al. Internet-delivered cognitive behavioral therapy to treat insomnia: a systematic review and meta-analysis. PLoS One 2016;11(2):e0149139 [FREE Full text] [doi: 10.1371/journal.pone.0149139] [Medline: 26867139]
11. Zachariae R, Lyby MS, Ritterband LM, O'Toole MS. Efficacy of internet-delivered cognitive-behavioral therapy for insomnia - A systematic review and meta-analysis of randomized controlled trials. Sleep Med Rev 2016 Dec;30:1-10. [doi: 10.1016/j.smrv.2015.10.004] [Medline: 26615572]
12. HealthTech Lab Inc. URL: https://www.htech-lab.co.jp/ [accessed 2021-10-26]
13. Bastien CH, Vallières A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. Sleep Med 2001 Jul;2(4):297-307. [Medline: 11438246]
14. Munezawa T. Development of the Japanese version of the Insomnia Severity Index (ISI-J). Japanese Journal of Psychiatric Treatment 2009;24:1-25.
15. Cella M, Chalder T. Measuring fatigue in clinical and community settings. J Psychosom Res 2010 Jul;69(1):17-22. [doi: 10.1016/j.jpsychores.2009.10.007] [Medline: 20630259]
16. World Health Organization Health and Work Performance Questionnaire (short form) Japanese edition. Research Institute of Occupational Mental Health. URL: http://riomh.umin.jp/lib/WHO-HPQ(Japanese).pdf [accessed 2021-10-26]

17. World Health Organization Health and Work Performance Questionnaire. Harvard Medical School. 2005. URL: https://www.hcp.med.harvard.edu/hpq/info.php [accessed 2022-06-17]
18. Frankel BL, Coursey RD, Buchbinder R, Snyder F. Recorded and reported sleep in chronic primary insomnia. Arch Gen Psychiatry 1976 May;33(5):615-623. [doi: 10.1001/archpsyc.1976.01770050067011] [Medline: 178287]
19. Hoch CC, Reynolds CF, Kupfer DJ, Berman SR, Houck PR, Stack JA. Empirical note: self-report versus recorded sleep in healthy seniors. Psychophysiology 1987 May;24(3):293-299. [doi: 10.1111/j.1469-8986.1987.tb00298.x] [Medline: 3602285]
20. Holmqvist M, Vincent N, Walsh K. Web- vs. telehealth-based delivery of cognitive behavioral therapy for insomnia: a randomized controlled trial. Sleep Med 2014 Feb;15(2):187-195. [doi: 10.1016/j.sleep.2013.10.013] [Medline: 24461370]
21. Kyle SD, Crawford MR, Morgan K, Spiegelhalder K, Clark AA, Espie CA. The Glasgow Sleep Impact Index (GSII): a novel patient-centred measure for assessing sleep-related quality of life impairment in insomnia disorder. Sleep Med 2013 Jun;14(6):493-501. [doi: 10.1016/j.sleep.2012.10.023] [Medline: 23347908]
22. Sharpe M, Wilks D. Fatigue. BMJ 2002 Aug 31;325(7362):480-483 [FREE Full text] [doi: 10.1136/bmj.325.7362.480] [Medline: 12202331]
23. Kapella MC, Herdegen JJ, Perlis ML, Shaver JL, Larson JL, Law JA, et al. Cognitive behavioral therapy for insomnia comorbid with COPD is feasible with preliminary evidence of positive sleep and fatigue effects. Int J Chron Obstruct Pulmon Dis 2011;6:625-635 [FREE Full text] [doi: 10.2147/COPD.S24858] [Medline: 22162648]
24. Heckler CE, Garland SN, Peoples AR, Perlis ML, Shayne M, Morrow GR, et al. Cognitive behavioral therapy for insomnia, but not armodafinil, improves fatigue in cancer survivors with insomnia: a randomized placebo-controlled trial. Support Care Cancer 2016 May;24(5):2059-2066 [FREE Full text] [doi: 10.1007/s00520-015-2996-y] [Medline: 26542272]

Abbreviations

BTI: brief therapy for insomnia
CBT-I: cognitive behavioral therapy for insomnia
CFS: Chalder fatigue scale
CONSORT: Consolidated Standards of Reporting Trials
HPQ: Health and Work Performance Questionnaire
ISI: Insomnia Severity Index
RCT: randomized controlled trial
SPA: sleep prompt app

Edited by R Kukafka; submitted 28.01.22; peer-reviewed by C Jacob, A Teles, Y Khan; comments to author 09.05.22; revised version received 10.06.22; accepted 23.06.22; published 25.07.22.

Please cite as:
Shimamoto T, Furihata R, Nakagami Y, Tateyama Y, Kobayashi D, Kiyohara K, Iwami T
Providing Brief Personalized Therapies for Insomnia Among Workers Using a Sleep Prompt App: Randomized Controlled Trial
J Med Internet Res 2022;24(7):e36862
URL: https://www.jmir.org/2022/7/e36862
doi:10.2196/36862
PMID:35877164

Original Paper

Impacts of Digital Care Programs for Musculoskeletal Conditions on Depression and Work Productivity: Longitudinal Cohort Study

Fabíola Costa[1], PhD; Dora Janela[1], PT; Maria Molinos[1], PhD; Robert Moulder[2], PhD; Vírgilio Bento[1], PhD; Jorge Lains[3,4], MD; Justin Scheer[5], MD; Vijay Yanamadala[1,6,7], MD, MBA; Steven Cohen[8,9], MD, PhD; Fernando Dias Correia[1,10], MD, PhD

[1]SWORD Health Inc, Draper, UT, United States
[2]Institute for Cognitive Science, University of Colorado, Boulder, CO, United States
[3]Rovisco Pais Medical and Rehabilitation Centre, Tocha, Portugal
[4]Faculty of Medicine, Coimbra University, Coimbra, Portugal
[5]Department of Neurological Surgery, University of California, San Francisco, CA, United States
[6]Department of Surgery, Frank H Netter School of Medicine, Quinnipiac University, Hamden, CT, United States
[7]Department of Neurosurgery, Hartford Healthcare Medical Group, Westport, CT, United States
[8]Department of Anesthesiology and Critical Care Medicine, Johns Hopkins School of Medicine, Baltimore, MD, United States
[9]Department of Physical Medicine and Rehabilitation, Uniformed Services University of the Health Sciences, Bethesda, MD, United States
[10]Neurology Department, Centro Hospitalar e Universitário do Porto, Porto, Portugal

Corresponding Author:
Fernando Dias Correia, MD, PhD
SWORD Health Inc
65 E Wadsworth Park Drive, Suite 230
Draper, UT, 84020
United States
Phone: 1 385 308 8034
Email: fcorreia@swordhealth.com

Abstract

Background: Comorbidity between musculoskeletal (MSK) pain and depression is highly common, and is associated with a greater symptom burden and greater loss of work productivity than either condition alone. Multimodal care programs tackling both physical and mental health components may maximize productivity recovery and return to work. Digital delivery of such programs can facilitate access, ensure continuity of care, and enhance patient engagement.

Objective: The aim of this study was to assess the impact of a completely remote multimodal digital care program (DCP) for MSK pain on mental health and work-related outcomes stratified by baseline depression levels.

Methods: Ad hoc analysis of an interventional, single-arm, cohort study of individuals with MSK pain undergoing a DCP was performed. Three subgroups with different baseline depression severity levels were established based on responses to the Patient Health Questionnaire (PHQ-9): cluster 1 (score<5: minimal depression), cluster 2 (scores 5-10: mild depression), and cluster 3 (score≥10: moderate depression). The mean changes in depression, anxiety, fear-avoidance beliefs, work productivity, and activity impairment and adherence between baseline and end of program (8-12 weeks) were assessed across subgroups by latent growth curve analysis.

Results: From a total of 7785 eligible participants, 6137 (78.83%) were included in cluster 1, 1158 (14.87%) in cluster 2, and 490 (6.29%) in cluster 3. Significant improvements in depression and anxiety scores were observed in clusters 2 and 3 but not in cluster 1, with average end-of-the program scores in clusters 2 and 3 below the initially defined cluster thresholds (score of 5 and 10, respectively). All clusters reported significant improvements in productivity impairment scores (mean changes from –16.82, 95% CI –20.32 to –13.42 in cluster 1 to –20.10, 95% CI –32.64 to –7.57 in cluster 3). Higher adherence was associated with higher improvements in depression in clusters 2 and 3, and with greater recovery in activities of daily living in cluster 3. Overall patient satisfaction was 8.59/10.0 (SD 1.74).

Conclusions: A multimodal DCP was able to promote improvements in productivity impairment scores comparable to those previously reported in the literature, even in participants with comorbid depression and anxiety. These results reinforce the need to follow a biopsychosocial framework to optimize outcomes in patients with MSK pain.

Trial Registration: ClinicalTrials.gov NCT04092946; https://clinicaltrials.gov/ct2/show/NCT04092946

(J Med Internet Res 2022;24(7):e38942) doi:10.2196/38942

KEYWORDS

musculoskeletal; pain; depression; anxiety; mental health; comorbidity; productivity; digital health; remote care; rehabilitation; telehealth; telemedicine; eHealth; digital health; digital care; multimodal; digital intervention; recovery; engagement; activities of daily living; work; job; occupational health; longitudinal cohort

Introduction

Musculoskeletal (MSK) pain is highly prevalent worldwide, affecting hundreds of millions of individuals with disability and personal suffering, while imposing a great socioeconomic burden [1]. Comorbidity between MSK pain and depression is very common [2] due to shared pathophysiological mechanisms, which establishes a strong and complex bidirectional relationship [3,4]. Comorbid depression symptoms have been associated with an increased MSK pain symptom burden and impaired recovery of both conditions [5,6]. This negative synergistic effect translates into poor work productivity [7,8], either by impaired performance at work (presenteeism) or work absence (absenteeism); decreased general quality of life; medical complications; and subsequent additional care [7-9]. Comorbid depression and MSK pain are associated with higher health care expenditures [10] estimated at US $13,000 annually in the United States, which is almost double that estimated for chronic pain alone [11]. In the United States, the annual cost associated with productivity loss from depression amounts to US $1150 per individual [12] and US $44 billion [13] for society, while the indirect costs associated with MSK pain are estimated at US $264 billion [14].

Exercise-based physical therapy is a first-choice intervention to address MSK pain [15-17]. Current guidelines advise addressing depression (as well as other cognitive and psychological factors) as part of MSK pain management [15,16,18] through a biopsychosocial approach [19,20], including pain education, psychoeducation, or even specifically cognitive behavioral therapy (CBT). This biopsychosocial approach has been increasingly applied and is naturally evolving with the optimization and digitalization of health care. With more than 62.5% of the global population now able to access the internet (according to Worldwide Digital Population estimates as of January 2022 [21]), digital interventions may offer highly scalable solutions to deliver evidence-based interdisciplinary interventions [22], thereby democratizing access and improving the continuity of care in cases where specially trained health care practitioners may not be readily available [23], and also promoting adherence to treatment by facilitating therapeutic alliance (defined as collaboration between therapeutic participants to foster healing) [24,25]. Digital interventions have therefore been explored for the treatment of depression and MSK diseases [26,27]. In 2021, the US Department of Health and Human Services' Substance Abuse and Mental Health Services Administration released an evidence-based resource guide system recommending the use of telehealth for people with serious mental health disorders such as depression, noting that the benefits of telehealth services in this context may extend beyond improvement in morbid psychological conditions, including chronic pain and pain-related disability [28].

To date, digital intervention studies have been focused on either pain and disability [29-31] or on mental health and pain [26,27,32,33], with only a few studies assessing the impact of either dimension (MSK pain and depression) on work-related productivity [27,34-36].

Previously, we reported a multimodal digital care program (DCP) that integrates physical therapy exercise-based management with a psychoeducational component, including CBT, that aims to encourage patients to develop self-management skills and strategies for their pain. This DCP has been validated in different MSK conditions in chronic [37], acute [38,39], and postsurgical contexts [40-43]. Herein, we aimed to assess mental health and work-related outcomes after a completely remote multimodal DCP for patients with MSK pain stratified by baseline depression levels.

We hypothesized that this multimodal DCP would be able to contribute to mental health and promote productivity impairment improvements, despite differences in the initial mental status of participants.

Methods

Study Design

This is an ad hoc analysis of a decentralized, single-arm ongoing study, focused on assessing clinical and engagement-related outcomes in patients with MSK pain after a home-based multimodal DCP.

Ethics Approval

This study was prospectively approved by the New England Institutional Review Board (number 120190313) and was registered on ClinicalTrials.gov (NCT04092946) on September 17, 2019.

Participants

Beneficiaries of employer health plans, older than 18 years of age, and suffering from MSK pain (either in the spine or in the upper or lower limbs) were invited to apply for SWORD Health's DCP through a dedicated website (which preselected candidates with ability to interact with technologies). Exclusion criteria were: (1) a health condition (eg, cardiac, respiratory) incompatible with at least 20 minutes of light to moderate exercise; (2) receiving active treatment for cancer; (3) reporting new-onset, rapidly progressive loss of strength and/or numbness in the arms/legs; or (4) reporting an unexplained change in

bowel or urinary function in the previous 2 weeks. Informed consent was obtained from all participants.

Intervention

The DCP was delivered between August 1, 2020, and October 12, 2021. This completely remote DCP integrates individually tailored exercises and a psychoeducational component, which includes both education and CBT. Upon enrollment, each participant was assigned to a physical therapist who was responsible for program customization and asynchronous monitoring of patient performance. The exercise sessions consist of gradual progressive movement exposure and are performed through a Food and Drug Administration–listed class II medical device, including a tablet with a preinstalled app and wearable motion-tracking sensors. The tablet displays the prescribed exercises through audio/videos, while sensors digitize motion, providing real-time biofeedback along with instructions to guide patients during their sessions. Data obtained from the exercise sessions are stored on a cloud-based platform, being asynchronously monitored through a web-based portal by the assigned physical therapist who adjusts the exercises according to the patient's progression. Participants were recommended to perform 3 exercise sessions per week, with an expected program duration ranging between 8 and 12 weeks depending on the condition (although early discharge was possible depending on physical therapist assessment). Absence of an exercise session for 28 consecutive days resulted in classification of the participant as a dropout. Participants were still considered if they were compliant with the intervention but failed to complete a given reassessment survey.

The psychoeducational component was developed under current clinical guidelines and research [17,44,45]. Educational articles were delivered through the app, covering a broad range of MSK pain–related topics, explaining pain and pain management. The CBT program consisted of self-guided interactive modules delivered through the smartphone app. This program was created by a multidisciplinary team including psychiatrists and psychologists based on third-generation CBT techniques, including mindfulness, acceptance and commitment therapy, and empathy-focused therapy. The CBT program was specifically designed to address fear avoidance, pain reconceptualization, active coping skills, as well as anxiety and depression associated with MSK pain. Bidirectional communication between participants and physical therapists, after exercise sessions or on demand, were ensured through a built-in secure chat feature on the smartphone app, and through synchronous video calls between the physical therapist and participant to facilitate therapeutic alliance, adjust treatment, and monitor potential adverse events.

Outcome Metrics

Self-reported assessments were collected at baseline, 4, 8, and 12 weeks, while mean changes were calculated between baseline and program end. Outcomes included mean change of (1) depression, measured by the 9-item Patient Health Questionnaire (PHQ-9; range 0-27) [46,47]; (2) anxiety, measured by the 7-item Generalized Anxiety Disorder (GAD-7) questionnaire (range 0-21) [48]; (3) Fear-Avoidance Beliefs Questionnaire for Physical Activity (FABQ-PA), comprising a total of 5 items each with a 7-option Likert scale (range 0-24) [49]; (4) Work Productivity and Activity Impairment (WPAI) questionnaire (0-100), including WPAI-overall (presenteeism and absenteeism from work), WPAI-work (presenteeism), WPAI-time (absenteeism, evaluated in employed participants), and WPAI-activity (for nonwork-related activity impairment in all participants) [50]; and (5) adherence, assessed through the number of completed sessions per week, total exercise time (minutes), communication frequency with the physical therapist, and overall satisfaction (points) through the question "On a scale from 0 to 10, how likely is it that you would recommend this intervention to a friend or neighbor? (from 0, not at all likely to 10, extremely likely). Overall, higher self-reported outcomes scores indicate higher severity.

Safety and Adverse Events

Participants were asked to grade the severity of pain and fatigue (from 0 to 10, with 10 being the most severe) on all exercise sessions to allow monitoring by the physical therapists. All communication channels were available for participants to report any adverse events. PHQ-9 and GAD-7 scores were used not only to guide the intervention approach but also to direct members to psychological and/or psychiatric care when needed, following the US Department of Health and Human Services guidelines.

Statistical Analysis

Study population demographics and clinical data, as well as usability metrics (number of sessions per week, total exercising time) were characterized through descriptive statistics. Differences in baseline characteristics between clusters (see below) were assessed through χ^2 tests for categorical variables and independent-samples *t* tests or one-way analysis of variance with Bonferroni posthoc correction for continuous variables.

Considering that depression has been reported as an important prognosis factor [51,52], PHQ-9 was used as a clustering variable, applying the thresholds <5 for no depression symptoms, 5-10 for mild symptoms, and ≥10 for moderate/severe symptoms, according to Kroenke et al [53].

For longitudinal data analysis, the latent growth curve was applied, which is a methodology in the same family as linear mixed-effects modeling, with the advantages of providing a measure of model fitness (eg, how well the model explains the data set) [54], and allowing the use of full information maximum likelihood (FIML) to address missing data. FIML has been shown to outperform other modern imputation models such as multiple imputation by chained equations (MICE) and listwise deletion [54-56]. FIML estimation considers all available data in each time point from all participants [55,56].

Latent growth curve analysis uses a structural equation model [57] to estimate the trajectories of outcomes over time based on individual trajectories and considering time as a continuous variable. This provides an estimate of the average trajectory (and respective pace of change) and individual variation around that trajectory over time (see Figure S1 in Multimedia Appendix 1).

The analysis was performed as an intent-to-treat analysis both for unfiltered cases and filtering for WPAI>0 points at baseline. Impact of training time on outcomes was modeled using cumulative training time as a time-invariant covariate. A conditional analysis was also performed to assess the influence of age, sex, and BMI as covariates. Models were adjusted for these covariates, which were fit as random effects allowing each to vary between individuals. All models were estimated with a robust sandwich estimator for standard errors.

Logistic regression was used to assess the relationship between baseline depression and productivity changes.

Significance levels were set at $P<.05$ in all analyses. Latent growth curve analysis was coded using R (version 1.4.1717) and all other analyses were performed using SPSS (version 17.0).

Results

Participant Characteristics

In total, 9388 participants were screened for eligibility, 621 (6.61%) of whom did not provide consent for research and 982 (10.46%) of whom were excluded (105 for clinical criteria, 37 participants missed their video call, and 840 declined to participate in the program during the video call) (Figure 1). In total, 7785 participants from all 50 states within the United States started the program. Overall, the completion rate was 77.05% (5998/7785), with 16.04% (962/5998), 19.02% (1141/5998), and 64.94% (3895/5998) participants discharged at 4, 8, and 12 weeks, respectively.

Using PHQ-9 as a clustering variable, 3 clusters were created: <5, 5-10, and ≥10 points, according to Kroenke et al [53]. Baseline demographics of each cluster are reported in Table 1 and those of the entire cohort are provided in Table S1 of Multimedia Appendix 2.

Figure 1. Flowchart of the study.

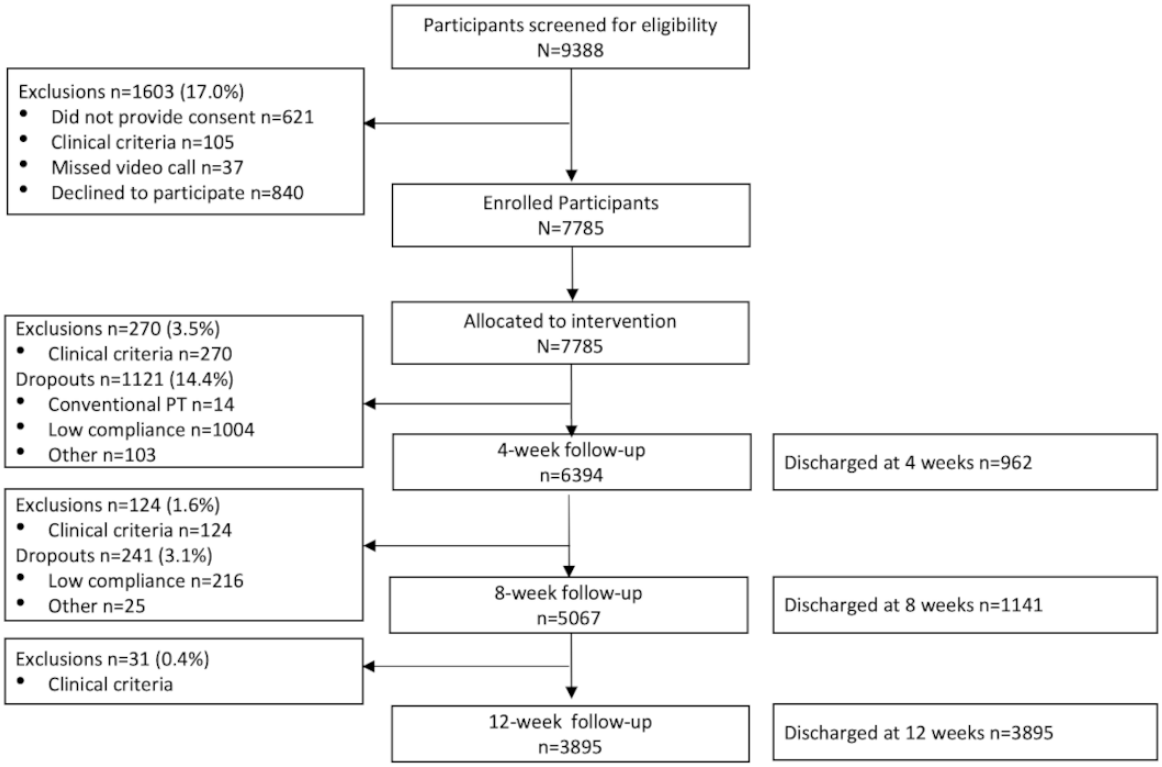


Table 1. Baseline characteristics of study participants in each depression-related cluster.

Characteristic	Cluster 1, PHQ-9[a] score <5 (n=6137)	Cluster 2, PHQ-9 score 5-10 (n=1158)	Cluster 3, PHQ-9 score ≥10 (n=490)	*P* value
Age (years), mean (SD)	51.4 (12.7)	50.0 (13.4)	48.7 (13.4)	<.001
Age category (years), n (%)				<.001
<25	43 (0.7)	16 (1.4)	8 (1.6)	
25-40	1339 (21.8)	298 (25.7)	146 (29.8)	
40-60	3172 (51.7)	561 (48.4)	225 (45.9)	
>60	1583 (25.8)	283 (24.4)	111 (22.7)	
Gender, n (%)				<.001
Woman	3276 (53.4)	702 (60.6)	297 (60.6)	
Man	2848 (46.4)	453 (39.1)	191 (39.0)	
Nonbinary	13 (0.2)	3 (0.3)	2 (0.4)	
BMI[b], mean (SD)	28.8 (6.3)	30.6 (7.1)	32.9 (7.9)	<.001
BMI category[b], n (%)				<.001
Underweight (<18.5)	46 (0.8)	10 (0.9)	3 (0.6)	
Normal (18.5-25)	1789 (29.2)	247 (21.4)	88 (18.0)	
Overweight (25-29)	2145 (35.0)	362 (31.3)	122 (25.0)	
Obese (30-39)	1775 (29.0)	412 (35.7)	197 (40.4)	
Obese grade III (>40)	367 (6.0)	124 (10.7)	78 (16.0)	
Conditions addressed, n (%)				<.001
Spine	2961 (48.2)	626 (54.1)	314 (64.1)	
Lower limb	1739 (28.3)	310 (26.8)	108 (22.0)	
Upper limb	1437 (23.4)	222 (19.2)	68 (13.9)	
Pain duration[c], n (%)				<.001
Acute (<12 weeks)	1620 (26.5)	235 (20.4)	68 (13.9)	
Chronic (>12 weeks)	4492 (73.5)	919 (79.6)	422 (86.1)	
Employment status, n (%)				<.001
Employed (part-time or full-time)	5260 (85.7)	992 (85.7)	364 (74.3)	
Unemployed	877 (14.3)	166 (14.3)	126 (25.7)	
Outcome measures, mean (SD)				
Pain level	4.67 (1.98)	5.10 (1.95)	5.73 (1.97)	<.001
Analgesics, n (%)	1861 (30.3)	464 (40.1)	212 (43.3)	<.001
Surgery intent	10.26 (19.52)	11.97 (21.41)	17.32 (26.86)	<.001
FABQ-PA[d,e]	10.27 (5.99)	11.63 (5.82)	12.72 (5.96)	<.001
GAD-7[f]	1.82 (2.88)	6.17 (4.23)	11.66 (5.62)	<.001
PHQ-9	0.86 (1.25)	6.93 (1.63)	14.84 (3.74)	<.001
WPAI[g]-overall[h]	14.78 (20.48)	23.94 (23.60)	40.87 (30.51)	<.001
WPAI-work[h]	13.86 (19.17)	22.08 (22.09)	38.04 (28.62)	<.001
WPAI-time[h]	1.98 (10.52)	4.36 (15.29)	12.02 (26.16)	<.001
WPAI-activity	25.43 (23.97)	34.47 (25.04)	50.49 (26.29)	<.001

[a]PHQ-9: 9-item Patient Health Questionnaire.

[b]n=20 missing values.

[c]n=29 missing values.

[d]FABQ-PA: Fear-Avoidance Beliefs Questionnaire for Physical Activity.

[e]n=10 missing values.

[f]GAD-7: 7-item Generalized Anxiety Disorder questionnaire.

[g]WPAI: Work Productivity and Activity Impairment questionnaire.

[h]N=6030.

Participants were unevenly distributed across the three clusters, with 6137, 1158, and 490 participants in cluster 1, 2, and 3, respectively. Several demographic characteristics were significantly different between clusters, namely the proportion of women, unemployed, younger participants, acuity, or those with higher BMI, which were progressively more frequent with increasing levels of depression (from cluster 1 to 2 and to 3, *P*<.001; Table 1). Significant baseline differences were also observed in clinical characteristics between clusters, as cluster 1 had no impairment in mental health (GAD-7 scores below 2 and FABQ 10.3, SD 6.0), and low levels of productivity impairment (WPAI-overall ~15%, SD 20.5 impairment). Cluster 2 reported mild anxiety and fear-avoidance beliefs (GAD-7 above 5 and FABQ 11.3, SD 5.8) and work difficulties (WPAI overall ~22%, SD 24.2 impairment), while cluster 3 presented with the highest anxiety (GAD-7 above 10) and greatest fear-avoidance beliefs (FABQ 12.7, SD 6.0) measures, along with greater impairments in productivity (40.87%, SD 30.5) or activities of daily living (50.5%, SD 26.3; *P*<.001; Table 1).

Clinical Outcomes

Overview

Changes in clinical outcomes over time are presented in Table 2, the model is presented in Table S2 of Multimedia Appendix 3, and the impact of covariates in the model is reported in Table S3 of Multimedia Appendix 4.

Table 2. Changes in clinical outcomes between baseline and end of program: intent-to-treat analysis.

Variable	Cluster 1			Cluster 2			Cluster 3		
	Baseline, mean (95% CI)	Change, mean (95% CI)	*P* value	Baseline, mean (95% CI)	Change, mean (95% CI)	*P* value	Baseline, mean (95% CI)	Change, mean (95% CI)	*P* value
PHQ-9[a]	0.81 (0.77 to 0.86)	–0.03 (–0.25 to 0.18)	.77	6.58 (6.44 to 6.71)	–3.36 (–4.26 to 2.45)	<.001	13.98 (13.46 to 14.5)	–4.09 (–6.87 to –1.32)	.004
GAD-7[b]	1.44 (1.35 to 1.53)	–0.43 (–0.68 to –0.18)	<.001	5.44 (5.07 to 5.8)	–2.25 (–3.14 to –1.36)	<.001	10.73 (9.99 to 11.47)	–2.24 (–4.49 to 0.01)	.05
FABQ-PA[c]	10.4 (10.18 to 10.62)	–2.63 (–3.17 to –2.09)	<.001	11.49 (10.98 to 12.01)	–2.33 (–3.62 to –1.04)	<.001	12.67 (11.92 to 13.42)	–0.7 (–2.84 to 1.44)	.52
WPAI[d]-overall	13.9 (13.11 to 14.7)	–6.84 (–8.82 to –4.86)	<.001	21.87 (19.39 to 24.35)	–11.33 (–16.6 to –6.06)	<.001	39.68 (34.86 to 44.5)	–12.34 (–23.65 to –1.03)	.03
WPAI-overall[e]	27.26 (26.1 to 28.41)	–16.82 (–20.23 to –13.42)	<.001	31.00 (28.26 to 33.74)	–19.11 (–25.75 to –12.47)	<.001	45.43 (40.62 to 50.24)	–20.1 (–32.64 to –7.57)	.002
WPAI-activity	22.57 (21.74 to 23.4)	–10.08 (–11.95 to –8.21)	<.001	30.99 (28.74 to 33.23)	–13.14 (–18.22 to –8.06)	<.001	46.35 (42.88 to 49.82)	–5.07 (–14.08 to 3.95)	.27
WPAI-activity[e]	32.04 (31.14 to 32.95)	–15.66 (–17.98 to –13.34)	<.001	36.28 (34.08 to 38.48)	–16.17 (–21.57 to –10.77)	<.001	49.5 (46.16 to 52.83)	–7.01 (–16.52 to 2.5)	.15
WPAI-work	13.04 (12.3 to 13.79)	–6.20 (–8.00 to –4.39)	<.001	20.78 (18.4 to 23.16)	–11.37 (–16.17 to –6.57)	<.001	37.13 (32.59 to 41.67)	–13.54 (–24.42 to –2.65)	.01
WPAI-work[e]	26.06 (24.98 to 27.13)	–15.24 (–18.35 to –12.12)	<.001	29.58 (26.93 to 32.24)	–19.36 (–25.29 to –13.42)	<.001	42.97 (38.44 to 47.51)	–21.9 (–34.43 to –9.37)	<.001
WPAI-time	2.02 (1.58 to 2.46)	–0.98 (–2.09 to 0.12)	.06	3.67 (2.01 to 5.34)	–1.42 (–3.93 to 1.09)	.23	13.45 (8.94 to 17.97)	–5.63 (–11.05 to –0.21)	.03
WPAI-time[e]	27.02 (22.52 to 31.52)	–28.54 (–39.11 to –17.96)	<.001	25.65 (17.24 to 34.06)	–20.5 (–31.84 to –9.16)	<.001	46.1 (35.5 to 56.7)	–23.13 (–43.29 to –2.98)	.02

[a]PHQ-9: 9-item Patient Health Questionnaire.

[b]GAD-7: 7-item Generalized Anxiety Disorder questionnaire.

[c]FABQ-PA: Fear-Avoidance Beliefs Questionnaire for Physical Activity.

[d]WPAI: Work Productivity and Activity Impairment questionnaire.

[e]Filtered for scores>0 at baseline.

Depression

Significant improvement was observed on the PHQ-9 for clusters 2 and 3, but not for cluster 1, which reported minimal depression symptoms at baseline (Table 2). Average end-of-the program scores for depression in clusters 2 and 3 decreased to levels below the initially defined lower cluster thresholds (5 and 10 for clusters 2 and 3, respectively; end scores: 3.22, 95% CI 2.31-4.13 and 9.89, 95% CI 7.15-12.62, respectively).

Significant differences in PHQ-9 score changes were observed between all clusters (P<.001, Table 3). Women reported higher levels of depression than men in cluster 1 (P<.001), but recovered at a faster pace in clusters 1 and 2. Older participants in cluster 3 recovered from depression symptoms at a slower pace (P=.04).

Table 3. Differences in clinical outcomes change stratified by baseline depression level.

Outcome	Cluster 1 versus cluster 2		Cluster 1 versus cluster 3		Cluster 2 versus cluster 3	
	Difference (95% CI)	*P* value	Difference (95% CI)	*P* value	Difference (95% CI)	*P* value
PHQ-9[a]	–2.44 (–3.37 to –1.51)	<.001	–9.11 (–11.85 to –6.36)	<.001	–6.66 (–9.55 to –3.78)	<.001
GAD-7[b]	–2.17 (–3.08 to –1.25)	<.001	–7.47 (–9.71 to –5.23)	<.001	–5.3 (–7.69 to –2.92)	<.001
FABQ-PA[c]	–1.39 (–2.79 to 0.01)	.05	–4.2 (–6.43 to –1.96)	<.001	–2.81 (–5.33 to –0.28)	.03
WPAI[d]-overall	–3.48 (–8.96 to 2.01)	.22	–20.28 (–31.72 to –8.84)	<.001	–16.8 (–29.21 to –4.4)	.008
WPAI-overall[e]	–2.29 (–5.17 to 9.75)	.55	–3.28 (–9.71 to –16.28)	.62	–1.00 (–13.19 to 15.18)	.07
WPAI-activity	–5.35 (–10.74 to 0.04)	.05	–28.79 (–37.88 to –19.7)	<.001	–23.43 (–33.69 to –13.18)	<.001
WPAI-activity[e]	–0.51 (–5.37 to 6.39)	.87	–8.65 (–18.44 to 1.14)	.08	–9.16 (–20.10 to 1.77)	.11
WPAI-work	–2.57 (–7.53 to 2.38)	.31	–16.75 (–27.85 to –5.64)	.003	–14.18 (–26.08 to –2.27)	.02
WPAI-work[e]	4.12 (–2.58 to 10.83)	.23	6.66 (–6.25 to 19.57)	.31	2.54 (–11.33 to 16.40)	.72
WPAI-time	–1.21 (–3.75 to 1.32)	.35	–6.78 (–12.53 to –1.03)	.02	–5.57 (–11.69 to 0.55)	.08
WPAI-time[e]	–8.04 (–23.54 to 7.47)	.31	–5.41 (–28.17 to 17.36)	.64	2.63 (–20.50 to 25.76)	.84

[a]PHQ-9: 9-item Patient Health Questionnaire.

[b]GAD-7: 7-item Generalized Anxiety Disorder questionnaire.

[c]FABQ-PA: Fear-Avoidance Beliefs Questionnaire for Physical Activity.

[d]WPAI: Work Productivity and Activity Impairment questionnaire.

[e]Filtered for scores>0 at baseline.

Anxiety

Cluster 1 did not report significant anxiety levels at baseline; thus, the observed change was not meaningful. Clusters 2 and 3 showed statistically significant improvements after the DCP (Table 3), ending the program with lower levels of anxiety than at baseline (end scores: 3.18, 95% CI 2.31-4.06 and 8.49, 95% CI 6.27-10.71, respectively).

Changes in anxiety levels were again significantly different among clusters (P<.001, Table 3). Women reported greater anxiety levels than men at baseline in clusters 1 and 2, whereas a faster-paced recovery was observed across all clusters for women (P<.001, Table S3 in Multimedia Appendix 4). Older participants in cluster 3 recovered from anxiety at a slower pace (P=.02). BMI did not influence mental health improvement trajectories in any cluster (Table S3 in Multimedia Appendix 4).

Fear-Avoidance Beliefs

Statistically significant improvements were observed in clusters 1 and 2, with a mean change of approximately –2.50 in both clusters (P<.001, Table 2). No significant improvement was observed in cluster 3 (P=.52, Table 2). BMI did not influence FABQ improvement trajectories in any cluster (Table S3 in Multimedia Appendix 4).

Work Productivity

Baseline impairments in work and in activities of daily living were progressively higher from cluster 1 to cluster 3 (Table 1). Across all WPAI scores, similar mean changes were observed between clusters (Table 3) when filtering for participants reporting impairments at baseline, despite different baseline values (Table 2).

Age and BMI did not consistently influence productivity impairment improvement across all clusters (Table S3 in Multimedia Appendix 4). Among women, we observed deacceleration in the improvement pace toward the end of the program for the WPAI overall score (Table S3 in Multimedia Appendix 4) as well as higher baseline scores paired with a slower recovery pace in WPAI activity.

Productivity impairment (WPAI-overall score) improved significantly across all clusters with mean changes ranging from –16.82 (95% CI –20.32 to –13.42) in cluster 1 and –19.11 (95% CI –25.75 to –12.47) in cluster 2 to –20.10 (95% CI –32.64 to –7.57) in cluster 3 when filtering for participants with reported

impairment at baseline (WPAI>0). A significant, albeit small, correlation between baseline PHQ-9 values and WPAI change was only observed in cluster 3 (0.30, *P*=.01).

Adherence and Usability-Related Outcomes

The average number of sessions per week was 2.7 (SD 1.39) across all clusters, but with individuals in clusters 2 and 3 showing slightly lower adherence (average of 2.6 sessions per week, SD 1.3; *P*<.001). This paralleled differences in the amount of time dedicated to exercise between clusters, which ranged from 552.4 minutes in cluster 1 to 384.7 minutes in cluster 3 (Table 4, *P*<.001). The influence of training time on outcome changes was estimated regarding trajectory slopes (Table 5).

In cluster 1, training time was significantly associated with FABQ reduction (*P*<.001) and improvement of WPAI-activity (*P*=.001). In cluster 2, increased training time was significantly associated with improvements in depression (*P*=.03) and FABQ (*P*=.002), but with no significant effect on productivity. In cluster 3, increased training times were significantly associated with greater improvements in mental health (depression [PHQ-9, *P*=.005], anxiety [GAD-7, *P*=.001], fear avoidance [FABQ, *P*=.003]) and in activities of daily living impairment (WPAI-activity, *P*<.001).

Regarding communication channels, the app chat was the preferred mode of contact, with an average of 9.1 days with contact. There were no significant differences between clusters (*P*=.21, Table 4), with cluster 3 showing a higher number of exchanged messages (Table 4). On average, 1.7 (SD 2.7) calls were made, which did not significantly differ between clusters (*P*=.22). Each participant on average engaged with 4.3 (SD 7.0) pieces of educational and CBT content, with no significant difference observed between clusters. Overall, the average satisfaction score was 8.6 (SD 1.7), which again did not significantly differ among the three clusters.

Table 4. Adherence outcome measures of the entire cohort and particular clusters.

Usability outcomes	Entire cohort, mean (SD)	Cluster 1, mean (SD)	Cluster 2, mean (SD)	Cluster 3, mean (SD)	*P* value
Number of sessions per week	2.7 (1.39)	2.7 (1.40)	2.6 (1.31)	2.6 (1.32)	<.001
Total exercising time, minutes	531.0 (522.3)	552.4 (537.4)	479.1 (473.6)	384.7 (392.6)	<.001
Average satisfaction (NRS[a], 0-10)	8.6 (1.74)[b]	8.6 (1.72)	8.5 (1.77)	8.6 (1.98)	.59
Number of days with contact (chat)	9.1 (11.6)	9.1 (11.6)	8.8 (11.5)	10.0 (11.7)	.21
Number of messages exchanged	23.74 (27.10)	23.30 (26.86)	24.37 (27.20)	28.37 (29.65)	.02
Number of educational articles per week	4.29 (7.0)	4.24 (7.0)	4.56 (6.9)	4.32 (7.2)	.37

[a]NRS: numerical rating scale.

Table 5. Effect of cumulative training time on the slopes of recovery trajectories for the different outcome variables.[a]

Variable	Cluster 1		Cluster 2		Cluster 3	
	Estimate	*P* value	Estimate	*P* value	Estimate	*P* value
PHQ-9[b]	–0.01	.18	–0.06	.03	–0.24	.005
GAD-7[c]	–0.01	.23	–0.05	.13	–0.20	.001
FABQ-PA[d]	–0.07	<.001	–0.13	.002	–0.21	.003
WPAI[e]-overall[f]	0.01	.92	0.13	.57	–0.56	.19
WPAI-activity[f]	–0.24	.001	–0.30	.10	–1.31	<.001
WPAI-work[f]	–0.03	.78	0.15	.48	–0.46	.29
WPAI-time[f]	0.39	.29	0.36	.34	–0.64	.49

[a]Negative values refer to more sharp slopes, indicating faster change over time.

[b]PHQ-9: 9-item Patient Health Questionnaire.

[c]GAD-7: 7-item Generalized Anxiety Disorder questionnaire.

[d]FABQ-PA: Fear-Avoidance Beliefs Questionnaire for Physical Activity.

[e]WPAI: Work Productivity and Activity Impairment questionnaire.

[f]Filtered for scores>0 at baseline.

Discussion

Principal Findings

Depression has been reported to be an important prognostic factor in MSK pain management [58,59]. By clustering patients according to baseline depression, we created three distinct groups, where other concomitant psychological factors (anxiety and fear-avoidance behaviors) with known prognostic value [5,60,61] were also observed at similar levels of severity. This is in accordance with the general population demographics, where depression is accompanied by other psychological/cognitive impairments [62,63], particularly in those with MSK pain [64]. Importantly, other demographic characteristics known to have prognostic value were distributed differently between clusters. Cluster 3 patients, besides suffering from more severe mental distress, were also more likely to be women [19], have a higher BMI [65], and be unemployed [66]. Cluster 2 contained these same factors at lower proportions but had a higher proportion of older participants [60], while cluster 1 tended to have more patients devoid of known risk factors.

Gender identification as a woman and older age were the predominant prognostic factors impacting both baseline and change in mental health status, while productivity was mainly influenced by gender identification as a woman. These results highlight the need to fine-tune programs to the specific needs of women, since this population is frequently identified as being more prone to psychological distress and prolonged MSK pain [67,68].

Notably, we observed that the improvements in mental health scores in clusters 2 and 3 resulted in average scores at the end of the program below the threshold used to define those clusters: cluster 2 from mild depression symptoms to minimal or no symptoms (<5) and cluster 3 from moderate to mild symptoms (<10). Overall, greater improvements were noted in cluster 3, demonstrating that multimodal DCP can not only impact physical health but also mental health in individuals with MSK pain and moderate mental health comorbidities.

Comparison to Prior Work

The observed improvements in mental health scores were higher than those previously reported by other multimodal telerehabilitation interventions [69,70], and were in line with those previously reported by us [37,39] and others [27,32,34,71]. Similar to the present study, Wang et al [71] reported a significant decrease in anxiety symptoms, allowing the transition to a lower level of anxiety according to established thresholds after a telerehabilitation program combining exercise with coaching.

A correlation between baseline depression and WPAI change was only observed in cluster 3, which might suggest that the prognostic value of depression may be dependent on higher severity stages [72,73] or that it is more relevant in chronic conditions [8,72,74]. However, it may also reflect that there is little room for improvement in psychological indices in patients with little baseline psychopathology.

Regarding productivity, all clusters had significant WPAI improvements, independent of the recoveries reported for fear-avoidance beliefs assessment, which in this study appeared to be a poor predictor for work-related productivity recovery. In fact, while fear-avoidance belief has been systematically associated with disability and pain [75-77], its correlation with work-related outcomes might only be observed in those with high FABQ or FABQ-work subscale scores [78]. Similar improvements in productivity were reported by Bailey et al [34] in a cohort study involving more than 10,000 participants with MSK pain, either with or without depression symptoms at baseline. Overall, the results are supportive of the application of multimodal/biopsychosocial approaches to address psychological (mal)adaptation to the pain condition [61,79,80] and maximize treatment outcomes, as previously reported with other telerehabilitation interventions [81,82].

Interestingly, the improvement observed for productivity was not replicated to the same extent in activities of daily living, particularly in cluster 3. The observed difference might be explained by the high level of depression symptoms and also by the greater prevalence of participants who were women, as highlighted by the conditional analysis, with women still being more frequently responsible for family and household activities [83,84].

The importance of patient compliance to obtain clinically meaningful outcomes is well-established [85,86]. We observed high compliance with exercise sessions across clusters, even in cluster 3 where, besides higher depression levels, >50% of participants were obese, a known factor for reduced engagement [65,87]. Herein, increased amounts of time dedicated to exercise sessions were associated with greater improvements in mental health and activities of daily living recovery, as previously reported [88-90].

However, we observed lower adherence to the psychoeducational program than anticipated, despite high adherence with other components of the program, high satisfaction levels, and frequent communication with the assigned physical therapist. The same challenges have been highlighted by other authors [91], suggesting that additional innovation to stimulate engagement in such components might further contribute to better outcomes [92-94]. Communication is key for establishing a therapeutic alliance (defined as collaboration between therapeutic participants to facilitate healing), which in turn is key for improved outcomes [60,95-97]. Digital interventions have been reported to promote similar or even better therapeutic alliance than in-person interventions [25]. Herein, communication between physical therapists and individuals was frequent, highlighting the convenience of the app chat. Further studies are needed to clarify the extent to which this accessible and available communication system impacts outcome.

Strengths and Limitations

Strengths of the study included the large sample size derived from a real-world context and the nature of the intervention: a multicomponent DCP managed by physical therapists combining exercises with real-time biofeedback within a biopsychosocial framework [17,95]. The digital format favors accessibility, while the regular communication with the same physical therapist may enhance adherence and maximize clinical outcomes [86,98].

The high adherence reported herein was objectively assessed to minimize social desirability response bias. Other strengths were the assessment of productivity through validated and widely used measures, and the inclusion of a heterogeneous cohort from geographically diverse states.

The major limitation refers to the study design that did not include a control group and thus does not allow us to determine the degree to which the various components of the program may have contributed to its overall reported changes, and whether all components benefit patients alike. Nevertheless, this study focused on an exploratory analysis of real-world data to support further research. Some variability was observed in terms of DCP participation and completion rates. However, the statistical methodologies chosen took into account the inclusion of real-world data, with missing data being handled through FIML, a method robust to attrition bias. This study included both emotional and cognitive outcomes; however, the inclusion of other psychosocial variables or tracking of mental-specific pharmacologic treatments could improve statistical models and further explain the variance, namely in the productivity measures.

Conclusions

A multimodal DCP was able to promote significant improvements in productivity, comparable to those previously reported in the literature, even in participants with comorbid depression and anxiety at baseline. These results reinforce the need to follow a biopsychosocial framework to optimize outcomes in patients with MSK pain to maximize return to work.

Acknowledgments

The authors acknowledge the team of physical therapists responsible for the management of participants. The study was funded by SWORD Health.

Data Availability

All relevant data are included in the article or available as supplementary material. The protocol, deidentified data, and analysis codes may be provided upon request to the corresponding author.

Authors' Contributions

All authors made a significant contribution to the work reported, as follows: Study concept and design: FDC, VB, JL, and SC; Data acquisition: MM; Statistical analysis: RM; Interpretation of data: FC, DJ, MM, SC, and FDC; Drafting the work: FC and DJ; Critical revision of the manuscript for important intellectual content: all authors; Final approval of the version to be published: all authors. The study sponsor, SWORD Health, was involved in study design, data collection, and interpretation and writing of the manuscript.

Conflicts of Interest

FC, DJ, MM, VB, VY, and FDC are employees at SWORD Health, the study sponsor. FDC, VY, and VB also hold equity from SWORD Health. RM is an independent scientific consultant who was responsible for the statistical analysis, while SC, JL, and JS are independent scientific/clinical consultants who were funded by SWORD Health in connection with the development and execution of this article.

Multimedia Appendix 1
Example path diagram for the latent growth curve (LGC) models used in the current study (Figure S1).
[DOCX File , 103 KB - jmir_v24i7e38942_app1.docx]

Multimedia Appendix 2
Table S1. Baseline characteristics of the entire cohort.
[DOCX File , 16 KB - jmir_v24i7e38942_app2.docx]

Multimedia Appendix 3
Table S2. Unconditional latent growth curve analysis: intent-to-treat analysis.
[DOCX File , 28 KB - jmir_v24i7e38942_app3.docx]

Multimedia Appendix 4
Table S3. Conditional growth-mixture modeling analysis: intent-to-treat analysis.
[DOCX File , 29 KB - jmir_v24i7e38942_app4.docx]

References

1. El-Tallawy SN, Nalamasu R, Salem GI, LeQuang JAK, Pergolizzi JV, Christo PJ. Management of musculoskeletal pain: an update with emphasis on chronic musculoskeletal pain. Pain Ther 2021 Jun;10(1):181-209 [FREE Full text] [doi: 10.1007/s40122-021-00235-2] [Medline: 33575952]
2. Daré LO, Bruand P, Gérard D, Marin B, Lameyre V, Boumédiène F, et al. Co-morbidities of mental disorders and chronic physical diseases in developing and emerging countries: a meta-analysis. BMC Public Health 2019 Mar 13;19(1):304 [FREE Full text] [doi: 10.1186/s12889-019-6623-6] [Medline: 30866883]
3. Crofford LJ. Psychological aspects of chronic musculoskeletal pain. Best Pract Res Clin Rheumatol 2015 Feb;29(1):147-155 [FREE Full text] [doi: 10.1016/j.berh.2015.04.027] [Medline: 26267008]
4. Chopra K, Arora V. An intricate relationship between pain and depression: clinical correlates, coactivation factors and therapeutic targets. Expert Opin Ther Targets 2014 Feb;18(2):159-176. [doi: 10.1517/14728222.2014.855720] [Medline: 24295272]
5. Wong W, Li M, Yung P, Leong H. The effect of psychological factors on pain, function and quality of life in patients with rotator cuff tendinopathy: A systematic review. Musculoskelet Sci Pract 2020 Jun;47:102173. [doi: 10.1016/j.msksp.2020.102173] [Medline: 32452391]
6. Puntillo F, Giglio M, Paladini A, Perchiazzi G, Viswanath O, Urits I, et al. Pathophysiology of musculoskeletal pain: a narrative review. Ther Adv Musculoskelet Dis 2021;13:1759720X21995067 [FREE Full text] [doi: 10.1177/1759720X21995067] [Medline: 33737965]
7. Cochrane A, Higgins NM, Rothwell C, Ashton J, Breen R, Corcoran O, et al. Work outcomes in patients who stay at work despite musculoskeletal pain. J Occup Rehabil 2018 Sep;28(3):559-567. [doi: 10.1007/s10926-017-9748-4] [Medline: 29236203]
8. Melkevik O, Clausen T, Pedersen J, Garde AH, Holtermann A, Rugulies R. Comorbid symptoms of depression and musculoskeletal pain and risk of long term sickness absence. BMC Public Health 2018 Aug 06;18(1):981 [FREE Full text] [doi: 10.1186/s12889-018-5740-y] [Medline: 30081870]
9. 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 07;380(9836):37-43 [FREE Full text] [doi: 10.1016/S0140-6736(12)60240-2] [Medline: 22579043]
10. Lentz TA, Harman JS, Marlow NM, Beneciuk JM, Fillingim RB, George SZ. Factors associated with persistently high-cost health care utilization for musculoskeletal pain. PLoS One 2019;14(11):e0225125 [FREE Full text] [doi: 10.1371/journal.pone.0225125] [Medline: 31710655]
11. Bilal J, Berlinberg A, Trost J, Riaz IB, Bhattacharjee S. The influence of depression on health care expenditures among adults with spondylosis, intervertebral disc disorders, and other back problems in the United States. Pain Med 2020 Feb 01;21(2):e45-e53. [doi: 10.1093/pm/pny223] [Medline: 30445578]
12. Allen D, Hines EW, Pazdernik V, Konecny LT, Breitenbach E. Four-year review of presenteeism data among employees of a large United States health care system: a retrospective prevalence study. Hum Resour Health 2018 Nov 09;16(1):59 [FREE Full text] [doi: 10.1186/s12960-018-0321-9] [Medline: 30413168]
13. Stewart WF, Ricci JA, Chee E, Hahn SR, Morganstein D. Cost of lost productive work time among US workers with depression. JAMA 2003 Jun 18;289(23):3135-3144. [doi: 10.1001/jama.289.23.3135] [Medline: 12813119]
14. Dieleman JL, Cao J, Chapin A, Chen C, Li Z, Liu A, et al. US health care spending by payer and health condition, 1996-2016. JAMA 2020 Mar 03;323(9):863-884 [FREE Full text] [doi: 10.1001/jama.2020.0734] [Medline: 32125402]
15. Kolasinski SL, Neogi T, Hochberg MC, Oatis C, Guyatt G, Block J, et al. 2019 American College of Rheumatology/Arthritis Foundation Guideline for the management of osteoarthritis of the hand, hip, and knee. Arthritis Care Res 2020 Feb;72(2):149-162. [doi: 10.1002/acr.24131] [Medline: 31908149]
16. National Guideline Centre. Low Back Pain and Sciatica in Over 16s: Assessment and Management. London, UK: National Institute for Health and Care Excellence; 2016.
17. Lin I, Wiles L, Waller R, Goucke R, Nagree Y, Gibberd M, et al. What does best practice care for musculoskeletal pain look like? Eleven consistent recommendations from high-quality clinical practice guidelines: systematic review. Br J Sports Med 2020 Jan 02;54(2):79-86. [doi: 10.1136/bjsports-2018-099878] [Medline: 30826805]
18. Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain 2008 Feb;9(2):105-121. [doi: 10.1016/j.jpain.2007.09.005] [Medline: 18055266]
19. Skúladóttir H, Sveinsdottir H, Holden JE, Gunnarsdóttir TJ, Halldorsdottir S, Björnsdottir A. Pain, sleep, and health-related quality of life after multidisciplinary intervention for chronic pain. Int J Environ Res Public Health 2021 Sep 28;18(19):10233 [FREE Full text] [doi: 10.3390/ijerph181910233] [Medline: 34639534]
20. Marin TJ, Van Eerd D, Irvin E, Couban R, Koes BW, Malmivaara A, et al. Multidisciplinary biopsychosocial rehabilitation for subacute low back pain. Cochrane Database Syst Rev 2017 Jun 28;6:CD002193 [FREE Full text] [doi: 10.1002/14651858.CD002193.pub2] [Medline: 28656659]
21. Kemp S. Digital 2022: Global overview report. Datareportal. 2022 Jan 26. URL: https://datareportal.com/reports/digital-2022-global-overview-report [accessed 2022-01-30]

22. Moshe I, Terhorst Y, Philippi P, Domhardt M, Cuijpers P, Cristea I, et al. Digital interventions for the treatment of depression: a meta-analytic review. Psychol Bull 2021 Aug;147(8):749-786. [doi: 10.1037/bul0000334] [Medline: 34898233]
23. Singla DR, Raviola G, Patel V. Scaling up psychological treatments for common mental disorders: a call to action. World Psychiatry 2018 Jun;17(2):226-227. [doi: 10.1002/wps.20532] [Medline: 29856556]
24. Fiani B, Siddiqi I, Lee SC, Dhillon L. Telerehabilitation: development, application, and need for increased usage in the COVID-19 era for patients with spinal pathology. Cureus 2020 Sep 21;12(9):e10563 [FREE Full text] [doi: 10.7759/cureus.10563] [Medline: 33101809]
25. Wehmann E, Köhnen M, Härter M, Liebherz S. Therapeutic alliance in technology-based interventions for the treatment of depression: systematic review. J Med Internet Res 2020 Jun 11;22(6):e17195 [FREE Full text] [doi: 10.2196/17195] [Medline: 32525484]
26. Baumeister H, Paganini S, Sander LB, Lin J, Schlicker S, Terhorst Y, et al. Effectiveness of a guided internet- and mobile-based intervention for patients with chronic back pain and depression (WARD-BP): a multicenter, pragmatic randomized controlled trial. Psychother Psychosom 2021;90(4):255-268. [doi: 10.1159/000511881] [Medline: 33321501]
27. Sander LB, Paganini S, Terhorst Y, Schlicker S, Lin J, Spanhel K, et al. Effectiveness of a guided web-based self-help intervention to prevent depression in patients with persistent back pain: the PROD-BP randomized clinical trial. JAMA Psychiatry 2020 Oct 01;77(10):1001-1011 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.1021] [Medline: 32459348]
28. Telehealth for the Treatment of Serious Mental Illness and Substance Use Disorders. URL: https://store.samhsa.gov/sites/default/files/SAMHSA_Digital_Download/PEP21-06-02-001.pdf [accessed 2021-01-30]
29. Jansen-Kosterink S, Huis in 't Veld R, Wever D, Hermens H, Vollenbroek-Hutten M. Introducing remote physical rehabilitation for patients with chronic disorders by means of telemedicine. Health Technol 2015 May 7;5(2):83-90. [doi: 10.1007/s12553-015-0111-5]
30. Pekyavas NO, Ergun N. Comparison of virtual reality exergaming and home exercise programs in patients with subacromial impingement syndrome and scapular dyskinesis: Short term effect. Acta Orthop Traumatol Turc 2017 May;51(3):238-242. [doi: 10.1016/j.aott.2017.03.008] [Medline: 28446376]
31. Kosterink SM, Huis in 't Veld RMHA, Cagnie B, Hasenbring M, Vollenbroek-Hutten MMR. The clinical effectiveness of a myofeedback-based teletreatment service in patients with non-specific neck and shoulder pain: a randomized controlled trial. J Telemed Telecare 2010;16(6):316-321. [doi: 10.1258/jtt.2010.006005] [Medline: 20798425]
32. Bossen D, Veenhof C, Van Beek KEC, Spreeuwenberg PM, Dekker J, De Bakker DH. Effectiveness of a web-based physical activity intervention in patients with knee and/or hip osteoarthritis: randomized controlled trial. J Med Internet Res 2013 Nov 22;15(11):e257 [FREE Full text] [doi: 10.2196/jmir.2662] [Medline: 24269911]
33. Garcia LM, Birckhead BJ, Krishnamurthy P, Sackman J, Mackey IG, Louis RG, et al. An 8-week self-administered at-home behavioral skills-based virtual reality program for chronic low back pain: double-blind, randomized, placebo-controlled trial conducted during COVID-19. J Med Internet Res 2021 Feb 22;23(2):e26292 [FREE Full text] [doi: 10.2196/26292] [Medline: 33484240]
34. Bailey JF, Agarwal V, Zheng P, Smuck M, Fredericson M, Kennedy DJ, et al. Digital care for chronic musculoskeletal pain: 10,000 participant longitudinal cohort study. J Med Internet Res 2020 May 11;22(5):e18250 [FREE Full text] [doi: 10.2196/18250] [Medline: 32208358]
35. Calner T, Nordin C, Eriksson MK, Nyberg L, Gard G, Michaelson P. Effects of a self-guided, web-based activity programme for patients with persistent musculoskeletal pain in primary healthcare: A randomized controlled trial. Eur J Pain 2017 Jul;21(6):1110-1120. [doi: 10.1002/ejp.1012] [Medline: 28464364]
36. Malliaras P, Cridland K, Hopmans R, Ashton S, Littlewood C, Page R, et al. Internet and telerehabilitation-delivered management of rotator cuff-related shoulder pain (INTEL Trial): randomized controlled pilot and feasibility trial. JMIR Mhealth Uhealth 2020 Nov 18;8(11):e24311 [FREE Full text] [doi: 10.2196/24311] [Medline: 33206059]
37. Janela D, Costa F, Molinos M, Moulder RG, Lains J, Francisco GE, et al. Asynchronous and tailored digital rehabilitation of chronic shoulder pain: a prospective longitudinal cohort study. J Pain Res 2022;15:53-66. [doi: 10.2147/JPR.S343308] [Medline: 35035234]
38. Correia FD, Molinos M, Neves C, Janela D, Carvalho D, Luis S, et al. Digital rehabilitation for acute ankle sprains: prospective longitudinal cohort study. JMIR Rehabil Assist Technol 2021 Sep 30;8(3):e31247 [FREE Full text] [doi: 10.2196/31247] [Medline: 34499038]
39. Costa F, Janela D, Molinos M, Lains J, Francisco GE, Bento V, et al. Telerehabilitation of acute musculoskeletal multi-disorders: prospective, single-arm, interventional study. BMC Musculoskelet Disord 2022 Jan 04;23(1):29 [FREE Full text] [doi: 10.1186/s12891-021-04891-5] [Medline: 34983488]
40. Correia FD, Nogueira A, Magalhães I, Guimarães J, Moreira M, Barradas I, et al. Home-based rehabilitation with a novel digital biofeedback system versus conventional in-person rehabilitation after total knee replacement: a feasibility study. Sci Rep 2018 Jul 26;8(1):11299. [doi: 10.1038/s41598-018-29668-0] [Medline: 30050087]
41. Correia FD, Nogueira A, Magalhães I, Guimarães J, Moreira M, Barradas I, et al. Medium-term outcomes of digital versus conventional home-based rehabilitation after total knee arthroplasty: prospective, parallel-group feasibility study. JMIR Rehabil Assist Technol 2019 Feb 28;6(1):e13111 [FREE Full text] [doi: 10.2196/13111] [Medline: 30816849]

42. Correia FD, Molinos M, Luís S, Carvalho D, Carvalho C, Costa P, et al. Digitally assisted versus conventional home-based rehabilitation after arthroscopic rotator cuff repair: a randomized controlled trial. Am J Phys Med Rehabil 2022 Mar 01;101(3):237-249 [FREE Full text] [doi: 10.1097/PHM.0000000000001780] [Medline: 33935152]
43. Dias Correia F, Nogueira A, Magalhães I, Guimarães J, Moreira M, Barradas I, et al. Digital versus conventional rehabilitation after total hip arthroplasty: a single-center, parallel-group pilot study. JMIR Rehabil Assist Technol 2019 Jun 21;6(1):e14523 [FREE Full text] [doi: 10.2196/14523] [Medline: 31228176]
44. Williams ACDC, Eccleston C, Morley S. Psychological therapies for the management of chronic pain (excluding headache) in adults. Cochrane Database Syst Rev 2012 Nov 14;11:CD007407 [FREE Full text] [doi: 10.1002/14651858.CD007407.pub3] [Medline: 23152245]
45. Joypaul S, Kelly F, McMillan SS, King MA. Multi-disciplinary interventions for chronic pain involving education: A systematic review. PLoS One 2019 Oct 2;14(10):e0223306 [FREE Full text] [doi: 10.1371/journal.pone.0223306] [Medline: 31577827]
46. Bijker L, Sleijser-Koehorst MLS, Coppieters MW, Cuijpers P, Scholten-Peeters GGM. Preferred self-administered questionnaires to assess depression, anxiety and somatization in people with musculoskeletal pain - a modified Delphi study. J Pain 2020;21(3-4):409-417 [FREE Full text] [doi: 10.1016/j.jpain.2019.08.006] [Medline: 31487562]
47. Costantini L, Pasquarella C, Odone A, Colucci ME, Costanza A, Serafini G, et al. Screening for depression in primary care with Patient Health Questionnaire-9 (PHQ-9): a systematic review. J Affect Disord 2021 Jan 15;279:473-483. [doi: 10.1016/j.jad.2020.09.131] [Medline: 33126078]
48. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
49. George SZ, Stryker SE. Fear-avoidance beliefs and clinical outcomes for patients seeking outpatient physical therapy for musculoskeletal pain conditions. J Orthop Sports Phys Ther 2011 Apr;41(4):249-259. [doi: 10.2519/jospt.2011.3488] [Medline: 21335927]
50. Ospina MB, Dennett L, Waye A, Jacobs P, Thompson AH. A systematic review of measurement properties of instruments assessing presenteeism. Am J Manag Care 2015 Feb 01;21(2):e171-e185 [FREE Full text] [Medline: 25880491]
51. Miles CL, Pincus T, Carnes D, Homer KE, Taylor SJC, Bremner SA, et al. Can we identify how programmes aimed at promoting self-management in musculoskeletal pain work and who benefits? A systematic review of sub-group analysis within RCTs. Eur J Pain 2011 Sep;15(8):775. [doi: 10.1016/j.ejpain.2011.01.016] [Medline: 21354838]
52. de Rooij A, Roorda LD, Otten RHJ, van der Leeden M, Dekker J, Steultjens MPM. Predictors of multidisciplinary treatment outcome in fibromyalgia:a systematic review. Disabil Rehabil 2013 Mar;35(6):437-449. [doi: 10.3109/09638288.2012.699582] [Medline: 22889312]
53. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. Psychiatric Annals 2002 Sep 01;32(9):509-515. [doi: 10.3928/0048-5713-20020901-06]
54. Iacobucci D. Structural equations modeling: fit indices, sample size, and advanced topics. J Consum Psychol 2010 Jan;20(1):90-98. [doi: 10.1016/j.jcps.2009.09.003] [Medline: 25904163]
55. Pfaffel A, Schober B, Spiel C. A comparison of three approaches to correct for direct and indirect range restrictions: a simulation study. Pract Assess Res Evaluation 2016;21:6. [doi: 10.7275/x4ep-fv42]
56. Xiao J, Bulut O. Evaluating the performances of missing data handling methods in ability estimation from sparse data. Educ Psychol Meas 2020 Oct;80(5):932-954 [FREE Full text] [doi: 10.1177/0013164420911136] [Medline: 32855565]
57. McNeish D, Matta T. Differentiating between mixed-effects and latent-curve approaches to growth modeling. Behav Res Methods 2018 Aug;50(4):1398-1414. [doi: 10.3758/s13428-017-0976-5] [Medline: 29067672]
58. Hooten WM. Chronic pain and mental health disorders: shared neural mechanisms, epidemiology, and treatment. Mayo Clin Proc 2016 Jul;91(7):955-970. [doi: 10.1016/j.mayocp.2016.04.029] [Medline: 27344405]
59. Kroenke K, Wu J, Bair MJ, Krebs EE, Damush TM, Tu W. Reciprocal relationship between pain and depression: a 12-month longitudinal analysis in primary care. J Pain 2011 Sep;12(9):964-973 [FREE Full text] [doi: 10.1016/j.jpain.2011.03.003] [Medline: 21680251]
60. Farin E, Gramm L, Schmidt E. The patient-physician relationship in patients with chronic low back pain as a predictor of outcomes after rehabilitation. J Behav Med 2013 Jun;36(3):246-258. [doi: 10.1007/s10865-012-9419-z] [Medline: 22476813]
61. Tseli E, Boersma K, Stålnacke BM, Enthoven P, Gerdle B, Äng BO, et al. Prognostic factors for physical functioning after multidisciplinary rehabilitation in patients with chronic musculoskeletal pain: a systematic review and meta-analysis. Clin J Pain 2019 Feb;35(2):148-173 [FREE Full text] [doi: 10.1097/AJP.0000000000000669] [Medline: 30371517]
62. Steffen A, Nübel J, Jacobi F, Bätzing J, Holstiege J. Mental and somatic comorbidity of depression: a comprehensive cross-sectional analysis of 202 diagnosis groups using German nationwide ambulatory claims data. BMC Psychiatry 2020 Mar 30;20(1):142 [FREE Full text] [doi: 10.1186/s12888-020-02546-8] [Medline: 32228541]
63. Thaipisuttikul P, Ittasakul P, Waleeprakhon P, Wisajun P, Jullagate S. Psychiatric comorbidities in patients with major depressive disorder. Neuropsychiatr Dis Treat 2014;10:2097-2103. [doi: 10.2147/NDT.S72026] [Medline: 25419132]
64. Simula AS, Ruokolainen O, Oura P, Lausmaa M, Holopainen R, Paukkunen M, et al. Association of STarT Back Tool and the short form of the Örebro Musculoskeletal Pain Screening Questionnaire with multidimensional risk factors. Sci Rep 2020 Jan 14;10(1):290. [doi: 10.1038/s41598-019-57105-3] [Medline: 31937867]

65. Vincent HK, Adams MCB, Vincent KR, Hurley RW. Musculoskeletal pain, fear avoidance behaviors, and functional decline in obesity: potential interventions to manage pain and maintain function. Reg Anesth Pain Med 2013;38(6):481-491. [doi: 10.1097/AAP.0000000000000013] [Medline: 24141874]
66. Ang DC, Bair MJ, Damush TM, Wu J, Tu W, Kroenke K. Predictors of pain outcomes in patients with chronic musculoskeletal pain co-morbid with depression: results from a randomized controlled trial. Pain Med 2010 Apr;11(4):482-491. [doi: 10.1111/j.1526-4637.2009.00759.x] [Medline: 20002592]
67. Wijnhoven HAH, de Vet HCW, Picavet HSJ. Prevalence of musculoskeletal disorders is systematically higher in women than in men. Clin J Pain 2006 Oct;22(8):717-724. [doi: 10.1097/01.ajp.0000210912.95664.53] [Medline: 16988568]
68. Narusyte J, Ropponen A, Mittendorfer-Rutz E, Svedberg P. Shared liability to pain, common mental disorders, and long-term work disability differs among women and men. Pain 2020 May;161(5):1005-1011 [FREE Full text] [doi: 10.1097/j.pain.0000000000001787] [Medline: 31977936]
69. Priebe JA, Haas KK, Moreno Sanchez LF, Schoefmann K, Utpadel-Fischler DA, Stockert P, et al. Digital treatment of back pain versus standard of care: the cluster-randomized controlled trial, Rise-uP. J Pain Res 2020;13:1823-1838. [doi: 10.2147/JPR.S260761] [Medline: 32765057]
70. Huber S, Priebe JA, Baumann K, Plidschun A, Schiessl C, Tölle TR. Treatment of low back pain with a digital multidisciplinary pain treatment app: short-term results. JMIR Rehabil Assist Technol 2017 Dec 04;4(2):e11 [FREE Full text] [doi: 10.2196/rehab.9032] [Medline: 29203460]
71. Wang G, Yang M, Hong M, Krauss J, Bailey JF. Clinical outcomes one year after a digital musculoskeletal (MSK) program: an observational, longitudinal study with nonparticipant comparison group. BMC Musculoskelet Disord 2022 Mar 11;23(1):237 [FREE Full text] [doi: 10.1186/s12891-022-05188-x] [Medline: 35277161]
72. Pinheiro MB, Ferreira ML, Refshauge K, Maher CG, Ordoñana JR, Andrade TB, et al. Symptoms of depression as a prognostic factor for low back pain: a systematic review. Spine J 2016 Jan 01;16(1):105-116. [doi: 10.1016/j.spinee.2015.10.037] [Medline: 26523965]
73. Orenius T, Koskela T, Koho P, Pohjolainen T, Kautiainen H, Haanpää M, et al. Anxiety and depression are independent predictors of quality of life of patients with chronic musculoskeletal pain. J Health Psychol 2013 Feb;18(2):167-175. [doi: 10.1177/1359105311434605] [Medline: 22453165]
74. Alhowimel A, AlOtaibi M, Radford K, Coulson N. Psychosocial factors associated with change in pain and disability outcomes in chronic low back pain patients treated by physiotherapist: A systematic review. SAGE Open Med 2018;6:2050312118757387 [FREE Full text] [doi: 10.1177/2050312118757387] [Medline: 29449945]
75. Smith JA, Russo L, Santayana N. Fear avoidance predicts persistent pain in young adults with low back pain: a prospective study. J Orthop Sports Phys Ther 2021 Aug;51(8):383-391 [FREE Full text] [doi: 10.2519/jospt.2021.9828] [Medline: 33998262]
76. Zale EL, Ditre JW. Pain-related fear, disability, and the fear-avoidance model of chronic pain. Curr Opin Psychol 2015 Oct 01;5:24-30 [FREE Full text] [doi: 10.1016/j.copsyc.2015.03.014] [Medline: 25844393]
77. Zale EL, Lange KL, Fields SA, Ditre JW. The relation between pain-related fear and disability: a meta-analysis. J Pain 2013 Oct;14(10):1019-1030 [FREE Full text] [doi: 10.1016/j.jpain.2013.05.005] [Medline: 23850095]
78. Wertli MM, Rasmussen-Barr E, Weiser S, Bachmann LM, Brunner F. The role of fear avoidance beliefs as a prognostic factor for outcome in patients with nonspecific low back pain: a systematic review. Spine J 2014 May 01;14(5):816-836. [doi: 10.1016/j.spinee.2013.09.036] [Medline: 24412032]
79. Chester R, Jerosch-Herold C, Lewis J, Shepstone L. Psychological factors are associated with the outcome of physiotherapy for people with shoulder pain: a multicentre longitudinal cohort study. Br J Sports Med 2018 Feb;52(4):269-275 [FREE Full text] [doi: 10.1136/bjsports-2016-096084] [Medline: 27445360]
80. Haik MN, Alburquerque-Sendín F, Fernandes RAS, Kamonseki DH, Almeida LA, Liebano RE, et al. Biopsychosocial aspects in individuals with acute and chronic rotator cuff related shoulder pain: classification based on a decision tree analysis. Diagnostics 2020 Nov 10;10(11):928 [FREE Full text] [doi: 10.3390/diagnostics10110928] [Medline: 33182699]
81. Blanquero J, Cortés-Vega MD, Rodríguez-Sánchez-Laulhé P, Corrales-Serra B, Gómez-Patricio E, Díaz-Matas N, et al. Feedback-guided exercises performed on a tablet touchscreen improve return to work, function, strength and healthcare usage more than an exercise program prescribed on paper for people with wrist, hand or finger injuries: a randomised trial. J Physiother 2020 Oct;66(4):236-242 [FREE Full text] [doi: 10.1016/j.jphys.2020.09.012] [Medline: 33069608]
82. Irvine AB, Russell H, Manocchia M, Mino DE, Cox Glassen T, Morgan R, et al. Mobile-web app to self-manage low back pain: randomized controlled trial. J Med Internet Res 2015 Jan 02;17(1):e1 [FREE Full text] [doi: 10.2196/jmir.3130] [Medline: 25565416]
83. Eek F, Axmon A. Gender inequality at home is associated with poorer health for women. Scand J Public Health 2015 Mar;43(2):176-182. [doi: 10.1177/1403494814562598] [Medline: 25504654]
84. Cerrato J, Cifre E. Gender inequality in household chores and work-family conflict. Front Psychol 2018;9:1330. [doi: 10.3389/fpsyg.2018.01330] [Medline: 30123153]
85. Nicolson PJA, Bennell KL, Dobson FL, Van Ginckel A, Holden MA, Hinman RS. Interventions to increase adherence to therapeutic exercise in older adults with low back pain and/or hip/knee osteoarthritis: a systematic review and meta-analysis. Br J Sports Med 2017 May;51(10):791-799. [doi: 10.1136/bjsports-2016-096458] [Medline: 28087567]

86. Jordan JL, Holden MA, Mason EE, Foster NE. Interventions to improve adherence to exercise for chronic musculoskeletal pain in adults. Cochrane Database Syst Rev 2010 Jan 20(1):CD005956 [FREE Full text] [doi: 10.1002/14651858.CD005956.pub2] [Medline: 20091582]
87. Samdal GB, Eide GE, Barth T, Williams G, Meland E. Effective behaviour change techniques for physical activity and healthy eating in overweight and obese adults; systematic review and meta-regression analyses. Int J Behav Nutr Phys Act 2017 Mar 28;14(1):42 [FREE Full text] [doi: 10.1186/s12966-017-0494-y] [Medline: 28351367]
88. Maestroni L, Read P, Bishop C, Papadopoulos K, Suchomel TJ, Comfort P, et al. The benefits of strength training on musculoskeletal system health: practical applications for interdisciplinary care. Sports Med 2020 Aug;50(8):1431-1450. [doi: 10.1007/s40279-020-01309-5] [Medline: 32564299]
89. Russell S, Jariwala A, Conlon R, Selfe J, Richards J, Walton M. A blinded, randomized, controlled trial assessing conservative management strategies for frozen shoulder. J Shoulder Elbow Surg 2014 Apr;23(4):500-507. [doi: 10.1016/j.jse.2013.12.026] [Medline: 24630545]
90. Hanratty CE, McVeigh JG, Kerr DP, Basford JR, Finch MB, Pendleton A, et al. The effectiveness of physiotherapy exercises in subacromial impingement syndrome: a systematic review and meta-analysis. Semin Arthritis Rheum 2012 Dec;42(3):297-316. [doi: 10.1016/j.semarthrit.2012.03.015] [Medline: 22607807]
91. Aragonès E, Rambla C, López-Cortacans G, Tomé-Pires C, Sánchez-Rodríguez E, Caballero A, et al. Effectiveness of a collaborative care intervention for managing major depression and chronic musculoskeletal pain in primary care: A cluster-randomised controlled trial. J Affect Disord 2019 Jun 01;252:221-229. [doi: 10.1016/j.jad.2019.04.004] [Medline: 30986737]
92. 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]
93. Bennell KL, Marshall CJ, Dobson F, Kasza J, Lonsdale C, Hinman RS. Does a web-based exercise programming system improve home exercise adherence for people with musculoskeletal conditions?: a randomized controlled trial. Am J Phys Med Rehabil 2019 Oct;98(10):850-858. [doi: 10.1097/PHM.0000000000001204] [Medline: 31021823]
94. Palazzo C, Klinger E, Dorner V, Kadri A, Thierry O, Boumenir Y, et al. Barriers to home-based exercise program adherence with chronic low back pain: Patient expectations regarding new technologies. Ann Phys Rehabil Med 2016 Apr;59(2):107-113 [FREE Full text] [doi: 10.1016/j.rehab.2016.01.009] [Medline: 27050664]
95. Keefe FJ, Main CJ, George SZ. Advancing psychologically informed practice for patients with persistent musculoskeletal pain: promise, pitfalls, and solutions. Phys Ther 2018 May 01;98(5):398-407 [FREE Full text] [doi: 10.1093/ptj/pzy024] [Medline: 29669084]
96. Hutting N, Caneiro JP, Ong'wen OM, Miciak M, Roberts L. Patient-centered care in musculoskeletal practice: Key elements to support clinicians to focus on the person. Musculoskelet Sci Pract 2022 Feb;57:102434 [FREE Full text] [doi: 10.1016/j.msksp.2021.102434] [Medline: 34376367]
97. Geraghty AWA, Roberts LC, Stanford R, Hill JC, Yoganantham D, Little P, et al. Exploring patients' experiences of internet-based self-management support for low back pain in primary care. Pain Med 2020 Sep 01;21(9):1806-1817 [FREE Full text] [doi: 10.1093/pm/pnz312] [Medline: 31841156]
98. Slade SC, Patel S, Underwood M, Keating JL. What are patient beliefs and perceptions about exercise for nonspecific chronic low back pain? A systematic review of qualitative studies. Clin J Pain 2014 Nov;30(11):995-1005. [doi: 10.1097/AJP.0000000000000044] [Medline: 24300225]

Abbreviations

CBT: cognitive behavioral therapy
DCP: digital care program
FABQ-PA: Fear-Avoidance Beliefs Questionnaire for Physical Activity
FIML: full information maximum likelihood
GAD-7: Generalized Anxiety Disorder 7-item questionnaire
MICE: multiple imputation by chained equations
MSK: musculoskeletal
PHQ-9: Patient Health 9-item questionnaire
WPAI: Work Productivity and Activity Impairment questionnaire

Edited by T Leung; submitted 22.04.22; peer-reviewed by A Costanza, E van der Velde, F Babatunde; comments to author 05.06.22; revised version received 09.06.22; accepted 16.06.22; published 25.07.22.

Please cite as:
Costa F, Janela D, Molinos M, Moulder R, Bento V, Lains J, Scheer J, Yanamadala V, Cohen S, Dias Correia F
Impacts of Digital Care Programs for Musculoskeletal Conditions on Depression and Work Productivity: Longitudinal Cohort Study
J Med Internet Res 2022;24(7):e38942
URL: https://www.jmir.org/2022/7/e38942
doi:10.2196/38942
PMID:35714099

Original Paper

Efficacy of Personalized Diabetes Self-care Using an Electronic Medical Record–Integrated Mobile App in Patients With Type 2 Diabetes: 6-Month Randomized Controlled Trial

Eun Young Lee[1*], MD, PhD; Seon-Ah Cha[2*], MD; Jae-Seung Yun[3], MD, PhD; Sun-Young Lim[4], MS; Jin-Hee Lee[4], PhD; Yu-Bae Ahn[3], MD, PhD; Kun-Ho Yoon[1], MD, PhD; Min Kyung Hyun[5], PhD; Seung-Hyun Ko[3], MD, PhD

[1]Division of Endocrinology and Metabolism, Department of Internal Medicine, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

[2]Division of Endocrinology and Metabolism, Department of Internal Medicine, Wonkwang University Sanbon Hospital, Gunpo, Republic of Korea

[3]Division of Endocrinology and Metabolism, Department of Internal Medicine, St. Vincent's Hospital, College of Medicine, The Catholic University of Korea, Suwon, Republic of Korea

[4]Catholic Institute of Smart Healthcare Center, The Catholic University of Korea, Seoul, Republic of Korea

[5]Department of Preventive Medicine, College of Korean Medicine, Dongguk University, Gyeongju, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Seung-Hyun Ko, MD, PhD
Division of Endocrinology and Metabolism, Department of Internal Medicine, St. Vincent's Hospital
College of Medicine
The Catholic University of Korea
93, Jungbu-daero, Paldal-gu
Suwon, 16247
Republic of Korea
Phone: 82 0312498155
Fax: 82 031 253 8898
Email: shkoh718@gmail.com

Abstract

Background: A system that combines technology and web-based coaching can help treat chronic conditions such as diabetes. However, the effectiveness of apps in mobile health (mHealth) interventions is inconclusive and unclear due to heterogeneous interventions and varying follow-up durations. In addition, randomized controlled trial data are limited, and long-term follow-up is lacking, especially for apps integrated into electronic medical records.

Objective: We aimed to assess the effect of an electronic medical record–integrated mobile app for personalized diabetes self-care, focusing on the self-monitoring of blood glucose and lifestyle modifications, on glycemic control in patients with type 2 diabetes mellitus.

Methods: In a 26-week, 3-arm, randomized, controlled, open-label, parallel group trial, patients with type 2 diabetes mellitus and a hemoglobin A_{1c} (HbA_{1c}) level of ≥7.5% were recruited. The mHealth intervention consisted of self-monitoring of blood glucose with the automatic transfer of glucose, diet, and physical activity counseling data (iCareD system). Participants were randomly assigned to the following three groups: usual care (UC), mobile diabetes self-care (MC), and MC with personalized, bidirectional feedback from physicians (MPC). The primary outcome was the change in HbA_{1c} levels at 26 weeks. In addition, diabetes-related self-efficacy, self-care activities, and satisfaction with the iCareD system were assessed after the intervention.

Results: A total of 269 participants were enrolled, and 234 patients (86.9%) remained in the study at 26 weeks. At 12 weeks after the intervention, the mean decline in HbA_{1c} levels was significantly different among the 3 groups (UC vs MC vs MPC: −0.49% vs −0.86% vs −1.04%; *P*=.02). The HbA_{1c} level decreased in all groups; however, it did not differ among groups after 26 weeks. In a subgroup analysis, HbA_{1c} levels showed a statistically significant decrease after the intervention in the MPC group compared with the change in the UC or MC group, especially in patients aged <65 years (*P*=.02), patients with a diabetes duration of ≥10 years (*P*=.02), patients with a BMI of ≥25.0 kg/m^2 (*P*=.004), patients with a C-peptide level of ≥0.6 ng/mL (*P*=.008), and

patients who did not undergo treatment with insulin (P=.004) at 12 weeks. A total of 87.2% (137/157) of the participants were satisfied with the iCareD system.

Conclusions: The mHealth intervention for diabetes self-care showed short-term efficacy in glycemic control, and the effect decreased over time. The participants were comfortable with using the iCareD system and exhibited high adherence.

Trial Registration: Clinical Research Information Service, Republic of Korea KCT0004128; https://tinyurl.com/bdd6pa9m

(J Med Internet Res 2022;24(7):e37430) doi:10.2196/37430

KEYWORDS

type 2 diabetes mellitus; digital health; mobile health; mHealth; mobile app; self-monitoring blood glucose; mobile phone

Introduction

Background

Diabetes is one of the most important chronic diseases that threatens public health [1]. Since 2000, the prevalence of diabetes has more than tripled, and by 2021, more than 530 million people worldwide will have diabetes [1]. The main goal of diabetes management is to maintain glycemic control within the target range, which is often accomplished through lifestyle modification and the self-monitoring of blood glucose (SMBG) in patients with type 2 diabetes mellitus (T2DM) [2,3]. However, maintaining glycemic control is challenging for both patients and health care providers (HCPs) because it is difficult to encourage or motivate patients to make long-term lifestyle changes, interpret their SMBG data, and provide immediate feedback and understand the patients' lifestyle due to brief clinic visit times and long visit intervals. Digital health care or mobile health (mHealth) services facilitate the collection of personal data, analyze data to evaluate clinical conditions, and provide personalized interventions or monitoring [4]. Through mHealth systems, patients with T2DM are encouraged to consume a healthy diet and perform physical activity (PA). Patient-reported data are used to tailor feedback messages, including health promotion, motivation, encouragement, reminders, and emotional support messages. Therefore, mHealth interventions may improve the health outcomes in patients with T2DM via tailored personalized interventions [5].

Recent mHealth interventions targeting patients with T2DM have diverse goals and components, including insulin-management apps, wearable blood glucose meters, automated text messages, health diaries, and virtual health coaching [6]. However, the effect of apps on mHealth interventions remains inconclusive and unclear because of heterogeneous interventions and various lengths of follow-up. On the basis of a mixed treatment comparison network meta-analysis using data from published randomized controlled trials (RCTs), mobile apps or apps with e-coaching interventions for patients with T2DM were more effective in improving the hemoglobin A_{1c} (HbA_{1c}) levels, fasting glucose, and hypoglycemia frequency than usual care (UC) during a 3- or 6-month follow-up period [7]. A meta-analysis of 13 studies on mobile apps for diabetes suggested overall efficacy in reducing HbA_{1c} levels, with a mean decrease of 0.44% (95% CI 0.29%-0.59%), as well as increased perception of self-care among mobile app users [8]. Various types of mHealth interventions have resulted in decrease in HbA_{1c} levels, in several RCTs that included patients with T2DM [9,10]. Most interventions demonstrated clinically and statistically significant efficacy, although some interventions had null results or achieved a <0.5% difference in the reduction of HbA_{1c} levels between the intervention and control groups [6].

Systems that combine technology and web-based coaching can be beneficial for treating diabetes and prediabetes [11,12]. To maintain lifestyle modification, patients should be continuously motivated and monitored in various ways, including individualized diabetes education and the use of aids based on information and communications technology [2,11]. Beyond using apps for personal use to collect and monitor lifelog data, web-based communication with physicians and HCPs would be more effective for diabetes self-care in patients with T2DM. Using this technology, physicians can view patient data in real time and between clinic visits and incorporate their lifelog data with clinical data derived from personal sensors and wearables in electronic medical records (EMRs). With regard to this type of intervention that connects a self-care app and EMRs, there is limited RCT data, and long-term follow-up is lacking.

Objectives

Therefore, we designed a diabetes management system using an EMR-integrated mobile app that provided regular feedback from HCPs to support diabetes self-care in a clinical setting for patients with T2DM. The app mainly offers lifestyle counseling to aid the SMBG, diet planning, and PA. The purpose of this study was to compare the clinical efficacy of a 26-week personalized diabetes self-care system using an EMR-integrated mobile app with that of UC in patients with T2DM. We also compared the effectiveness of this system with and without feedback from the HCPs.

Methods

Study Design

This study was designed as a 26-week open-label, parallel group, 3-arm RCT conducted in 2 separate university-affiliated hospitals from August 2019 to December 2021. A detailed description of the study design has been previously reported [13]. As shown in Figure 1, all participants were randomly assigned to 1 of the following 3 groups: group 1, UC; group 2, mobile diabetes self-care (MC); and group 3, MC with personalized, bidirectional feedback from physicians (MPC). Among the 279 patients screened, 269 (96.4%) were enrolled in the study, and of these, 234 (86.9%) completed the

intervention. A total of 269 individuals were included in the intention-to-treat (ITT) analyses (Figure 2).

Figure 1. Study design. HbA_{1c}: hemoglobin A_{1c}.

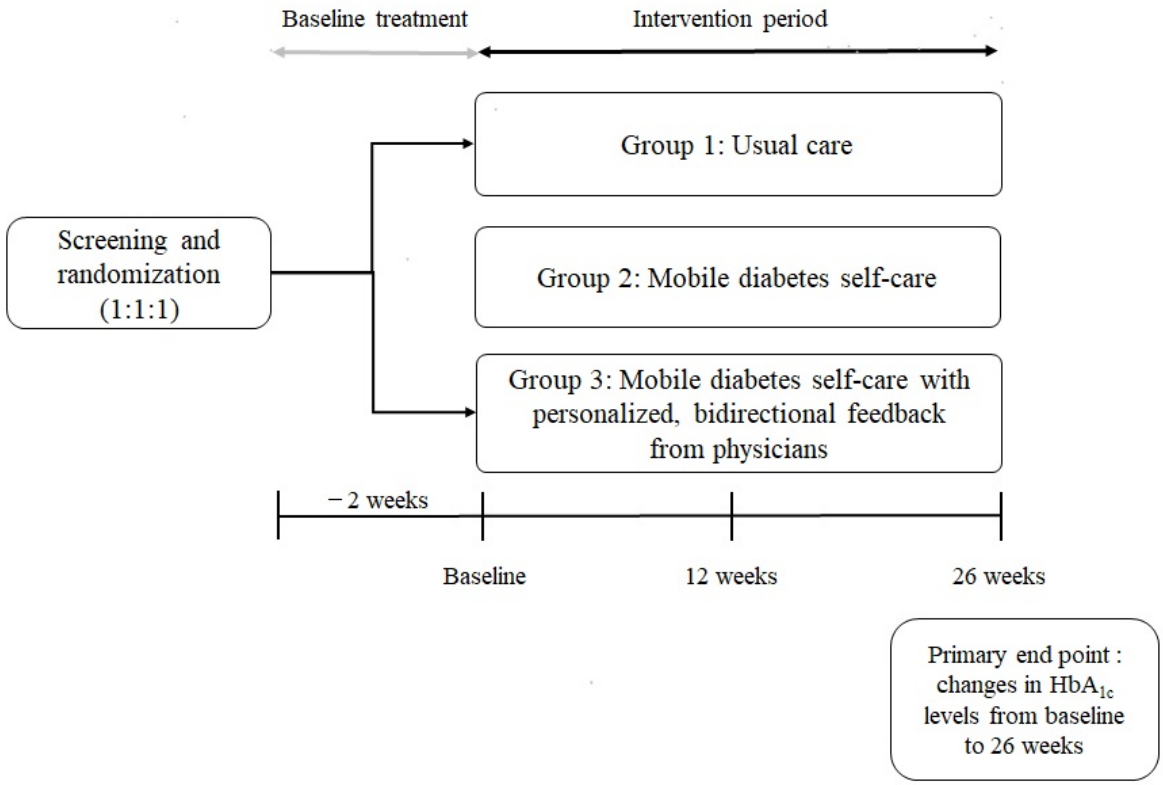
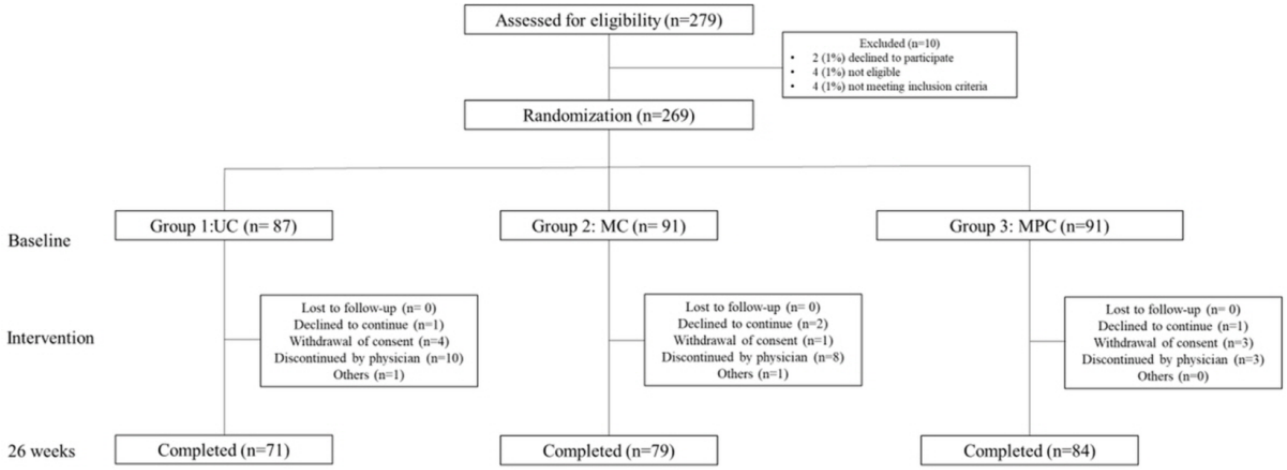


Figure 2. Flowchart of patient enrollment and status. MC: mobile diabetes self-care; MPC: mobile diabetes self-care with personalized, bidirectional feedback from physicians; UC: usual care.



Eligibility

Patients aged 19 to approximately 74 years with T2DM, an HbA_{1c} level of ≥7.5%, and a BMI of ≥18.5 kg/m^2 who could use a smartphone and consented to participate were eligible for this study. Participants who were using insulin pumps; who were pregnant; who had serious medical illness including end-stage renal disease, heart failure, and cancer; or who had difficulty performing PA were excluded. Further detailed inclusion and exclusion criteria are described in the study protocol [13]. The trial was registered with the Clinical Research Information Service, Republic of Korea (KCT0004128).

Sample Size

On the basis of previous studies [10,14,15], we assumed a mean difference in HbA_{1c} levels of at least 0.60 between the control and intervention groups and an SD, within groups, of 0.75 after 6 months. In a single-factor ANOVA study, 73 patients per group resulted in 90% power and a .05 significance level. Finally, 282 participants (94 patients per group) were required to achieve 73 patients per group after accounting for a predicted dropout rate of up to 20%.

Randomization

Those who signed the informed consent form were randomly assigned to 1 of 3 groups in a 1:1:1 ratio. Random numbers

were generated using SAS (version 9.3; SAS Institute) [13]. Stratification by institution and a baseline HbA_{1c} level of 8.5% were performed using the stratified permutated block randomization method. As this study had an open-label design, HCPs and participants were informed of the group assignment at the time of randomization. The participants and HCPs could not be blinded in this study because of the intervention method.

Intervention

Regardless of the assigned group, all participants were provided a glucometer (CareSens N; i-SENS, Inc) from which SMBG data were automatically transferred to our mHealth system, called the *iCareD system*. Basic diabetes education, including information on the SMBG, diet, and PA, was provided to all participants. The control group (UC) received UC according to the standard care for patients with T2DM by the Korean Diabetes Association [2]. Participants were instructed to perform the SMBG 4 times a day (before a meal in the morning and 2 hours after every meal), record their glucose levels in a notebook, and bring the notebook to their clinic visits. In the MC and MPC intervention groups, a diabetes self-care mobile app (iCareD; Medical Excellence Inc) was used in addition to the UC for diabetes management (Table S1 in Multimedia Appendix 1) [13]. The app allowed patients to enter their self-care data (SMBG, dietary habits, and step count), and automated text messages (educational, behavioral, and motivational messages) from the iCareD system were sent to their mobile phones. The automated messages were sent 3 times per week and consisted of 2 standardized messages for diabetes self-care and lifestyle modification and a customized message according to the lifestyle questionnaire. In the iCareD system, specific content for each message was developed based on the clinical practice guidelines of the Korean Diabetes Association and multidisciplinary expert opinions from our diabetes care team (endocrinologist, certified dietician, and diabetes educator). In both the MC and MPC groups, the mobile app was integrated with the EMR in each hospital; therefore, HCPs also evaluated participants at every 3-month visit based on the data obtained from the mobile app. Participants were instructed to upload their diet photos through the app. To encourage PA, we set the goal of a step count of >10,000 steps per day, and this goal was adjusted according to underlying diseases or individual health conditions. These data were also transferred to the iCareD system, which was integrated with the EMR system for HCPs in the hospital.

For the MPC group, based on our previous study, an HCP sent additional personalized recommendations and bidirectional feedback to each participant every 2 weeks through the iCareD system during the intervention period; the feedback was mainly related to diabetes self-care, the SMBG, or lifestyle modification [16].

In the offline system, patients visited the outpatient clinic every 3 months, and at these visits, physicians conducted face-to-face interviews with their patients, reviewed their uploaded data linked to the EMR, and provided individualized interventions based on these data. All participants were allowed to contact educator nurses over telephone but were encouraged to use the app.

Primary and Secondary Outcomes

The primary outcome was the difference in the change in HbA_{1c} levels (%) between baseline and 26 weeks among the 3 groups. The secondary outcome was the changes in HbA_{1c} and fasting glucose (mg/dL) levels between the UC and 2 mobile-based intervention groups between baseline and 26 weeks. The HbA_{1c} level <7% attainment rates were evaluated at 12 and 26 weeks. In addition, lifestyle changes based on PA and diet records; cardiometabolic risk factors such as body weight, blood pressure, and lipid profile; program satisfaction and compliance (or adherence); frequency of hypoglycemia; and changes in homeostasis model assessment of insulin resistance and β cell function were assessed at 26 weeks. Adherence was defined as the proportion of intervention participation using the iCareD app, including blood glucose measurement and feedback confirmation, over a 24-week period. Exploratory assessment variables included changes in diabetes prescriptions, SMBG frequency, and BMI.

Participant satisfaction was assessed in the 2 intervention groups by using a locally developed satisfaction survey at 26 weeks. The survey included 5-level Likert-type questions evaluating self-care efficacy and various opinions on the iCareD system, such as the ease and frequency of text messages, perceived efficacy, and willingness to continue with or recommend the iCareD program to family or friends. A score of 5 indicated *very satisfied* or *strongly agree*. Higher scores on the satisfaction scale reflect better results.

Measurements

Demographic and clinical information collected at baseline and follow-up has been described previously [13]. PA was tracked using a Google Fit mobile app and assessed as the total step count per day [17]. Body composition data were obtained using a bioimpedance analyzer (InBody 720 and 970, InBody Co, Ltd) at baseline and every 26 weeks. Laboratory parameters, including fasting glucose, HbA_{1c} level, and lipid profile, were collected at every visit. C-peptide and urinary albumin to creatinine ratios were measured at baseline and every 26 weeks. We used the updated homeostasis model assessment calculator to evaluate the homeostasis model assessment of insulin resistance and β cell function [18-20].

Hypoglycemic events, including hospitalization or emergency room visits due to hypoglycemia, blood glucose levels <70 mg/dL, or related symptoms even without the SMBG, were evaluated at every visit. Diabetes management behaviors such as SMBG frequency, PA, and diet records were obtained at every visit. SMBG frequency was defined as the average number of tests performed per day, calculated for each patient based on the records in the web system. The goal achievement rate for PA was defined as the number of days the target was reached/total measured days × 100 (%). User satisfaction with mobile app was surveyed in the MC and MPC groups.

Statistical Analysis

Continuous variables were presented as mean (SD), whereas categorical data were presented as frequencies with percentages. Analysis of covariance was used to compare the mean 26-week

HbA_{1c} levels among the 3 groups. Post hoc analysis was performed using the Bonferroni method. The number of hypoglycemic events among the groups was compared using the chi-square test or Fisher exact test. The goal achievement rate for PA was analyzed, except for the case of <1000 steps per day. Missing data were replaced by the last-observation-carried-forward method for all participants who were followed up at least once after enrollment. Both per-protocol and ITT analyses were conducted. Unless otherwise specified, analyses were performed based on the results of the ITT analysis. The analysis was performed using SAS (version 9.3; SAS Institute Inc). Statistical significance was set at *P* value of <.05.

Ethics Approval

The study protocol was approved by the ethics committee of St. Vincent's Hospital (VC19EEDI0085) and St. Mary's Hospital (KC19EEDE0278). All participants provided written informed consent before enrollment in the study. All data and information were anonymized according to the International Conference on Harmonization Good Clinical Practice guidelines.

Results

Participant Flow

During the recruitment period from August 2019 to August 2020 in the outpatient clinics of 2 separate university-affiliated diabetes centers, a total of 279 participants were assessed for eligibility and 269 (96.4%) participants were randomized. A total of 10 participants withdrew consent, leaving 269 participants to be included in this study (Figure 2).

After the 26-week follow-up, the total retention rate was 86.9% (234/269), with an equal distribution among the groups. The baseline analysis revealed no significant differences between those who completed the study and those who were lost to follow-up (data not shown).

Clinical Characteristics of Participants

The mean age of the participants was 52.5 (12.3) years, and 42.8% (115/269) of the participants were male. The mean baseline HbA_{1c} level and duration of diabetes were 8.7% (1.3%) and 11.4 (8.1) years, respectively. The mean BMI was 27.2 (4.6) kg/m^2, and 41.3% (111/269) of the participants had hypertension. None of the other baseline characteristics or variables differed significantly among the 3 study groups. There was no significant difference in the presence of microvascular and macrovascular complications between the 3 groups (all *P*>.05; Table 1).

Table 1. Baseline demographic and clinical characteristics of patients.

	Group 1: UC[a] (n=87)	Group 2: MC[b] (n=91)	Group 3: MPC[c] (n=91)	*P* value
Age (years), mean (SD)	52.6 (12.1)	51.3 (13.1)	53.6 (11.7)	.66
Sex (male), n (%)	37 (43)	40 (44)	38 (42)	.96
Duration of diabetes (years), mean (SD)	11.5 (8.2)	10.9 (8.3)	11.9 (7.8)	.61
Body weight (kg), mean (SD)	73.5 (17.2)	73.2 (16.9)	71.8 (13.0)	.97
BMI (kg/m^2), mean (SD)	27.4 (4.9)	27.3 (5.0)	26.8 (3.9)	.79
SBP[d] (mm Hg), mean (SD)	129.6 (14.5)	126.7 (15.0)	129.0 (14.5)	.38
DBP[e] (mm Hg), mean (SD)	78.0 (10.5)	77.1 (9.8)	76.8 (10.5)	.47
Current smokers, n (%)	10 (12)	11 (12)	6 (7)	.40
Alcohol consumption, n (%)	27 (31)	25 (28)	21 (23)	.49
Physical activity (step count per day), mean (SD)	6280.9 (3159.4)	6208.9 (3212.8)	6630.0 (3639.8)	.78
Education, n (%)				.25
Elementary	10 (12)	5 (6)	14 (15)	
High school	48 (55)	50 (55)	43 (47)	
College	29 (33)	36 (40)	34 (37)	
Comorbidities, n (%)				
Hypertension (yes)	39 (52)	36 (45)	36 (25)	.49
Hyperlipidemia (yes)	66 (88)	72 (90)	75 (89)	.92
Complication, n (%)				
CVD[f]	13 (15)	13 (14)	15 (17)	.91
Retinopathy	15 (17)	15 (17)	21 (23)	.46
Nephropathy	27 (31)	32 (35)	37 (41)	.40
Neuropathy	9 (10)	10 (11)	15 (17)	.40
Antidiabetic medications, n (%)				.38
Insulin only	1 (1)	3 (3)	2 (2)	
Oral agents	53 (61)	50 (55)	46 (51)	
Insulin+oral agents	23 (26)	28 (31)	35 (39)	
Laboratory measurements, mean (SD)				
Fasting glucose (mg/dL)	164.7 (50.0)	163.6 (60.8)	166.3 (61.2)	.28
eGFR[g] (ml/min/1.73 m^2)	101.3 (23.3)	99.5 (22.9)	95.0 (28.4)	.08
HbA_{1c}[h] level (%)	8.6 (1.1)	8.7 (1.3)	8.8 (1.4)	.78
Total cholesterol (mg/dL)	154.7 (37.0)	170.2 (58.5)	157.7 (37.8)	.16
Triglyceride (mg/dL)	158.1 (107.0)	177.1 (254.0)	156.5 (111.4)	.72
HDL[i] cholesterol (mg/dL)	48.2 (12.4)	50.3 (12.6)	48.2 (11.7)	.39
LDL[j] cholesterol (mg/dL)	76.2 (32.4)	86.6 (35.6)	79.4 (29.9)	.13
C-peptide level (ng/mL)	2.1 (1.2)	2.2 (1.6)	2.1 (1.3)	.93
HOMA-IR[k]	1.9 (1.1)	1.9 (1.5)	1.9 (1.2)	.86
HOMA-β[l]	49.7 (32.7)	53.0 (36.2)	51.5 (38.0)	.74

[a]UC: usual care.

[b]MC: mobile diabetes self-care.

[c]MPC: mobile diabetes self-care with personalized, bidirectional feedback from physicians.

[d]SBP: systolic blood pressure.
[e]DBP: diastolic blood pressure.
[f]CVD: cardiovascular disease.
[g]eGFR: estimated glomerular filtration rate.
[h]HbA_{1c}: hemoglobin A_{1c}.
[i]HDL: high-density lipoprotein.
[j]LDL: low-density lipoprotein.
[k]HOMA-IR: homeostasis model assessment for insulin resistance.
[l]HOMA-β: homeostasis model assessment for β-cell function.

Primary Outcome Measure: Change in HbA_{1c} Level

The change in HbA_{1c} levels did not differ significantly at 26 weeks among the 3 groups (Figure 3). However, the reduction in HbA_{1c} levels at 12 weeks was significantly different among the 3 groups (*P*=.02; Table 2). In the post hoc analysis, only the MPC group showed a significant decrease in HbA_{1c} levels compared with the UC group. (Table 2). In a subgroup analysis, a decrease in HbA_{1c} levels among the 3 groups showed a significant difference only at 12 weeks, especially in the patients aged <65 years (*P*=.02), patients with a diabetes duration of ≥10 years (*P*=.02), patients with a BMI of ≥25.0 kg/m^2 (*P*=.004), patients with a C-peptide level of ≥0.6 ng/mL (*P*=.008), and patients who did not undergo treatment with insulin (*P*=.004; Table S2 in Multimedia Appendix 2). Adjusting for age, sex, and baseline HbA_{1c} level did not affect the HbA_{1c} level change results.

Figure 3. Mean HbA_{1c} level from baseline to 26 weeks. HbA_{1c}: hemoglobin A_{1c}; MC: mobile diabetes self-care; MPC: mobile diabetes self-care with personalized, bidirectional feedback from physicians; UC: usual care.

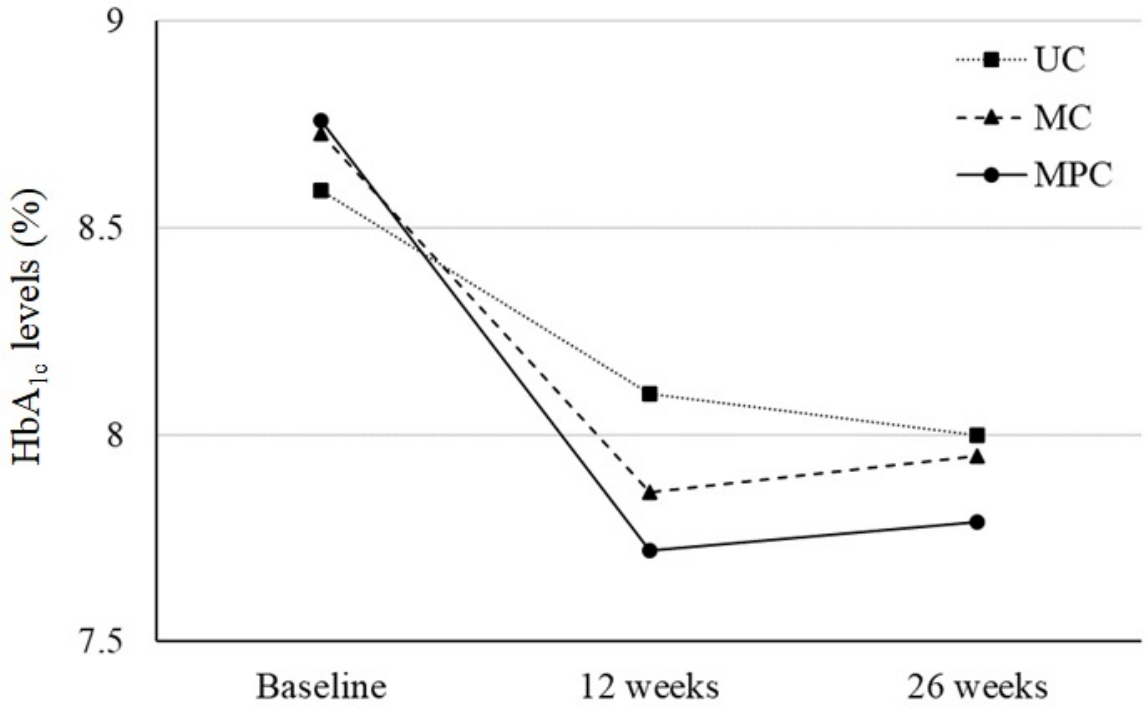


Table 2. Changes in hemoglobin A_{1c} (HbA_{1c}) level from baseline and HbA_{1c} level <7% attainment rate.

	Group 1: UC[a] (n=87)	Group 2: MC[b] (n=91)	Group 3: MPC[c] (n=91)	*P* value
Changes in HbA_{1c} level from baseline (%), mean (SD)				
Change at 12 weeks	−0.5 (1.0)	−0.9 (1.4)	−1.0 (1.5)[d]	.02
Change at 26 weeks	−0.6 (1.1)	−0.8 (1.7)	−1.0 (1.5)	.30
HbA_{1c} level <7% attainment rate, n (%)				
At 12 weeks	12 (14)	15 (17)	21 (23)	.25
At 24 weeks	15 (17)	15 (17)	26 (29)	.08

[a]UC: usual care.

[b]MC: mobile diabetes self-care.

[c]MPC: mobile diabetes self-care with personalized, bidirectional feedback from physicians.

[d]*P*<.05 versus group 1 in post hoc analysis.

Secondary Outcome Measures

The HbA_{1c} level <7% attainment rate increased from between the UC and MPC groups at 12 weeks, but the difference was not significant (UC vs MC vs MPC: 12/87, 14% vs 15/91, 17% vs 21/91, 23%; *P*=.25). The MPC group tended to have higher attainment rate of HbA_{1c} level <7% at 26 weeks, compared with the other groups (UC vs MC vs MPC: 15/87, 17% vs 15/91, 17% vs 26/91, 29%; *P*=.08; Table 2). Other changes in clinical and behavioral outcomes from baseline to follow-up are shown in Table 3. Fasting glucose levels were significantly reduced from baseline to each time point in all 3 groups (*P*<.05 for each group, each follow-up time). However, the change in fasting glucose levels did not differ among the 3 groups during the 26-week intervention period (Table 3). Changes in body weight and BMI from baseline to 26 weeks also showed no differences among the study groups.

The frequency of the SMBG did not show any significant differences among the 3 groups at the 26-week follow-up. However, compared with patients in the UC group, those in the 2 intervention groups (iCareD system users) tended to have more frequent SMBG recordings at 12 weeks (UC vs iCareD system users: 1.4 [1.1] times per day vs 1.6 [1.0] times per day; *P*=.09). The average frequency of the SMBG showed a negative correlation with HbA_{1c} (*r*=0.277; *P*=.003).

PA, defined as step counts per day, was not significantly different among the study groups 26 weeks after the intervention. The goal achievement rate for PA was higher in the MC and MPC groups than that in the UC group at 26 weeks, but the difference was not significant (UC vs MC vs MPC: 15.1% vs 18.5% vs 17.9%; *P*=.45). The changes in low-density lipoprotein–cholesterol level showed significant differences among the 3 groups at both 12 and 26 weeks (*P*=.001 for 12 weeks and *P*=.02 for 26 weeks). Low-density lipoprotein–cholesterol levels increased in the UC group and decreased in the MC and MPC groups during the follow-up period.

Table 3. Secondary study outcomes at baseline and follow-up.

	Group 1: UC[a] (n=87)	Group 2: MC[b] (n=91)	Group 3: MPC[c] (n=91)	*P* value
Fasting glucose (mg/dL), mean (SD)				
Baseline	164.7 (50.0)	163.6 (60.8)	166.3 (61.2)	.83
12 weeks	154.6 (54.7)	145.3 (51.7)	146.3 (45.3)	.28
26 weeks	148.9 (55.7)	146.3 (56.2)	142.7 (47.5)	.76
Change from baseline at 12 weeks	−10.1 (50.9)	−18.3 (64.3)	−20.0 (50.6)	.68
Change from baseline at 26 weeks	−15.8 (57.0)	−17.3 (64.5)	−23.7 (57.6)	.89
PA[d] (step counts/day), mean (SD)				
Baseline to approximately 12 weeks	6069.3 (2774.4)	6143.0 (2849.6)	6447.0 (3338.2)	.87
12 weeks to 26 weeks	5827.0 (2879.9)	6019.3 (2953.1)	6319.1 (3652.4)	.94
Body weight (kg), mean (SD)				
Baseline	73.5 (17.2)	73.2 (16.9)	71.9 (13.0)	.97
12 weeks	73.2 (17.3)	73.0 (17.1)	71.8 (13.2)	.98
26 weeks	73.1 (17.1)	73.2 (17.4)	71.8 (13.2)	.97
Change from baseline at 12 weeks	−0.27 (1.85)	−0.19 (2.05)	−0.07 (2.17)	.93
Change from baseline at 26 weeks	−0.42 (2.84)	0.03 (3.22)	−0.03 (3.51)	.98
BMI (kg/m^2), mean (SD)				
Baseline	27.4 (4.8)	27.3 (5.0)	26.9 (3.9)	.80
12 weeks	27.3 (4.8)	27.2 (5.1)	26.8 (4.0)	.84
26 weeks	27.3 (4.7)	27.3 (5.2)	26.8 (4.0)	.91
Change from baseline at 12 weeks	−0.11 (0.73)	−0.07 (0.72)	−0.03 (0.82)	.94
Change from baseline at 26 weeks	−0.18 (1.08)	0.01 (1.19)	−0.01 (1.33)	.97
LDL[e] cholesterol (mg/dL), mean (SD)				
Baseline	76.2 (32.4)	86.6 (35.6)	79.4 (29.9)	.13
12 weeks	80.5 (36.0)	80.3 (34.6)	76.2 (33.8)	.52
26 weeks	78.6 (33.0)	77.5 (33.7)	76.8 (34.7)	.86
Change from baseline at 12 weeks	4.8 (19.3)	−6.4 (29.2)	−5.1 (25.8)	.001
Change from baseline at 26 weeks	3.5 (18.7)	−9.1 (32.0)	−4.8 (30.3)	.02
HOMA-IR[f], mean (SD)				
Baseline	1.9 (1.1)	1.9 (1.5)	1.9 (1.2)	.88
26 weeks	1.8 (1.0)	1.9 (1.5)	1.8 (1.4)	.76
Change from baseline at 26 weeks	−0.1 (0.8)	0.0 (0.1)	−0.2 (1.2)	.81
HOMA-β[g], mean (SD)				
Baseline	49.7 (32.7)	53.0 (36.2)	51.5 (38.0)	.74
26 weeks	58.5 (30.2)	60.5 (35.3)	67.7 (80.6)	.99
Change from baseline at 26 weeks	8.6 (28.2)	6.9 (37.5)	14.5 (75.5)	.78

[a]UC: usual care.

[b]MC: mobile diabetes self-care.

[c]MPC: mobile diabetes self-care with personalized, bidirectional feedback from physicians.

[d]PA: physical activity.

[e]LDL: low-density lipoprotein.

[f]HOMA-IR: homeostasis model assessment for insulin resistance.

[g]HOMA-β: homeostasis model assessment for β-cell function.

Satisfaction for iCareD System

A total of 157 out of the 182 (86.2%) participants completed the satisfaction survey at 26 weeks of the intervention. Participants' satisfaction with the iCareD program was very high in both the MC and MPC groups. Overall, most patients were satisfied with the system (104/157, 66.2% strongly agree; 33/157, 21% agree), understood all the messages (27/157, 17% strongly agree; 89/157, 57% agree), were willing to use the program (56/157, 36% strongly agree; 73/157, 46% agree), felt that it helped them reach their goals (74/157, 47% strongly agree; 61/157, 39% agree), and recommended the iCareD system to a family member or friend with T2DM (57/157, 36% strongly agree; 77/157, 49% agree). No differences were observed between the 2 groups (Table 4).

Table 4. Messages read for 6 months (intervention period) and program satisfaction.

	Group 2: MC[a] (n=91)	Group 3: MPC[b] (n=91)	*P* value
Automated message sent, N	141	155	N/A[c]
Automated message read, n (%)	110 (78)	134 (86.4)	.09
Personalized message sent, N	N/A	12	N/A
Personalized message read, n (%)	N/A	12 (100)	<.001
Satisfaction survey			
Total, N	77	80	N/A
Overall satisfaction, n (%)	67 (87)	70 (88)	.39
Help to diabetes self-care, n (%)	67 (87)	69 (86)	.37
Will you recommend the app to others? n (%)	67 (87)	67 (84)	.92
Do you want to continue using the app? n (%)	63 (81)	66 (83)	.70

[a]MC: mobile diabetes self-care.

[b]MPC: mobile diabetes self-care with personalized, bidirectional feedback from physicians.

[c]N/A: not applicable.

There were no statistically significant differences between the MC and MPC groups in skill and technique acquisition, health service navigation, or manipulation of app content. iCareD system users especially valued convenient SMBG data reporting via automatic wireless transfer from the glucometer to the app without needing to write directly in the notebook and the ability to browse accumulated personal data (63/157, 40% strongly agree; 69/157, 44% agree). An end-of-intervention usability survey demonstrated that participants were comfortable with using the iCareD system.

With regard to adherence, compared with participants in the MC group, those in the MPC group checked the automated text messages from the iCareD system for 26 weeks. The proportion of participants who read >75% of the automated messages (3 times per week) was significantly higher in the MPC group than in the MC group (76/91, 83% vs 61/91, 67%; *P*=.02). The number of participants who uploaded photos of a meal and step count was higher in the MPC group than in the MC group, but the proportion declined to 50% by the end of the study in both groups.

Adverse Events

No serious adverse events were reported from enrollment until the completion of this study. Hypoglycemic events were infrequent and showed no differences among the groups at 26 weeks (Table 5). No deaths, direct study-related adverse events, or severe hypoglycemic episodes were reported or detected.

Table 5. Hypoglycemic events.

	Group 1: UC[a]	Group 2: MC[b]	Group 3: MPC[c]	*P* value
Patients who experienced hypoglycemia, N	87	91	91	N/A[d]
Baseline to <12 weeks, n (%)	15 (17)	34 (35)	32 (35)	.03
12 weeks to 26 weeks, n (%)	22 (25)	28 (34)	26 (30)	.84
Frequency of hypoglycemia[e], mean (SD)				
Baseline to <12 weeks	6.1 (7.4)	2.9 (3.1)	4.3 (5.6)	.25
12 weeks to 26 weeks	3.2 (4.1)	3.3 (2.9)	3.6 (3.8)	.56
Unexpected clinic visit or hospitalization due to hypoglycemia[f], n (%)	0 (0)	2 (2)	0 (0)	.18

[a]UC: usual care.

[b]MC: mobile diabetes self-care.

[c]MPC: mobile diabetes self-care with personalized, bidirectional feedback from physicians.

[d]N/A: not applicable.

[e]Per patient who experienced hypoglycemia for 90 days.

[f]Hypoglycemia was defined as a blood glucose level of <70 mg/dL.

Discussion

Principal Findings

This study was an RCT investigating a hospital-based, EMR-integrated mobile app–based diabetes self-care intervention over a 26-week period in patients with T2DM. Although the HbA_{1c} level decreased from the baseline value in all 3 groups, the HbA_{1c} changes did not show any significant difference between the control and 2 intervention groups at 26 weeks. However, we found that interactive mHealth intervention for diabetes self-care (MPC group) significantly decreased HbA_{1c} levels by −1.04% compared with −0.49% in the UC group and −0.86% in the MC group at 12 weeks. This finding was consistent among non–insulin users, those aged < 65 years, and those who were obese.

Owing to the global growth in the use of mobile phones with powerful platforms to help health care, many types of apps have been developed. According to Liquid-State, in 2018, there were >318,000 mHealth care apps available for patients, and approximately 200 new health care apps were being built each day [21]. In all, 70% of mHealth practitioners have reported that diabetes is currently the leading health target for the mobile app industry [22]. In 2017, more than 1500 diabetes-related apps were reported to be available to users.

Mobile apps related to diabetes management generally deal with information about diabetes, healthy diet, PA, weight loss, the SMBG, adherence, and motivation [6]. mHealth interventions support self-care and diabetes education and encourage lifestyle modification. These data may be used to tailor feedback messages or advice on specific behavior changes to implement; these messages are usually sent automatically according to an algorithm [5,9,23]. Compared with conventional mobile apps that collect only patient-driven data, our EMR-integrated mobile app could provide important clues to the future direction of mobile app development for diabetes management in 2 respects. First, HCPs can be provided with patients' medical history such as comorbidities or current medications, in addition to patient-centered data. Given the high rates of comorbidity and concurrent medications in patients with T2DM [24], this integrated provision of medical information may allow HCPs to provide accurate guidance to patients on diet, exercise, and management of comorbid diseases rather than simply focusing on the message to lower blood glucose levels. Second, from the patients' perspective, it is possible to provide better insight into diabetes management by providing laboratory results by time course along with personal data collection information. In particular, our systems adopted visualization of glucose levels by color to improve awareness or alertness of hyperglycemia (red) or hypoglycemia (black) [13].

Using the EMR-integrated mobile app intervention, we demonstrated a significant reduction in HbA_{1c} levels after 12 weeks of intervention. Consistent with our results, a 3-month RCT using DialBetics, a smartphone-based self-management support system for Japanese patients with T2DM, demonstrated that HbA_{1c} levels decreased by an average of 0.4% compared with an increase of 0.1% in the control group, with improvement in fasting glucose level and BMI [25]. A systematic review also revealed limited robust evidence of the promising short-term effectiveness of mHealth interventions for diabetes, such as the improvement of HbA_{1c} levels [14,15,26,27]. However, a caveat of these RCT analyses is that most of them included only studies conducted under highly controlled conditions with a small number of patients [14,15,26,27]. Interestingly, the first Food and Drug Administration–approved mobile app, BlueStar, showed no intervention effects in a real-world setting with >100 patients, despite significant reductions in HbA_{1c} (>1%) in their first RCT with 30 patients [28,29]. It is noteworthy that our study showed significant differences in HbA_{1c} levels among groups at 3 months in real-world practice, with a relatively large number of patients at 2 different clinical sites. This finding suggests the potential usefulness of EMR-integrated mobile app interventions in diabetes management. In addition, we found that the intervention effects in the MPC group were prominent in patients with younger age, obesity, higher C-peptide levels, and no insulin treatment. This finding implies that mobile-based

interventions, such as other diabetes treatments, may be more effective when β cell function is preserved. This also highlights the importance of early intervention. However, the intervention effect was also pronounced in those with a diabetes duration ≥10 years. This indicates that although early intervention may be important, such interventions may also be effective in long-standing diabetes.

Mobile phone apps that receive blood glucose data from a connected glucometer are available and have the capacity to make data upload and review less burdensome [30]. The internet-based SMBG system, which augments the SMBG by giving patients the means to communicate their blood glucose levels to their HCP for actional feedback, has been shown to reduce HbA_{1c} levels in some RCTs involving patients with T2DM [31]. The inverse correlation between reporting frequency and HbA_{1c} levels, as well as the significant difference in HbA_{1c} levels only for frequent testers (defined as those who test on average twice or more per day), suggests that frequent SMBG has an effect on reducing HbA_{1c} levels only when combined with regular, frequent communication of SMBG with an HCP [32]. The recording of a food diary using a smartphone app is a well-known simple tool, and technology to use images to quantify the composition and calorie content of food has been developed. However, it is difficult and cumbersome for users to constantly record data based on their eating habits [33,34]. However, automated integration of glucose and lifelog data in the EMR between scheduled clinic visits improves the HCP workflow for reviewing data and improves communication with patients, eventually leading to better care [35].

In this study, 87.2% (137/152) of the participants were satisfied with the iCareD system and answered that the app helped their diabetes self-care skills; however, the iCareD system failed to decrease HbA_{1c} levels over >3 months. There are possible explanations for the lack of improvement in HbA_{1c} levels in mHealth app users. First, age is a barrier to digital health care adoption and may influence the adoption of new technologies [36]. The mean age of the participants in this study was 52.5 years (range 20-74 years). A total of 16% (42/269) of the participants were aged >65 years. Second, the iCareD system was developed with a focus on lifestyle changes rather than strict glucose control or active medication adjustment, such as whirlwind dosage escalation of antidiabetic medications. In the case of the TExT-MED study, a unidirectional text message intervention for diabetes self-care providing text message triggers to encourage individuals to engage in self-care behaviors, the TExT-MED program also did not result in a significant improvement in HbA_{1c} levels. However, trends toward improvement in the primary outcome of HbA_{1c} levels and other secondary outcomes, including quality of life, were observed. Similar to our satisfaction survey, 94% (44/47) of the patients who received the TExT-MED intervention enjoyed the program and believed it was a good way to learn about diabetes [5]. Patient engagement was highest for more medical topics, such as glucose monitoring and medications, and lower for lifestyle topics, such as PA and healthy coping [6]. Therefore, we suggest that interventions for diabetes self-care should include improving HbA_{1c} levels through modification of lifestyle, glucose monitoring, and adherence and dosage adjustment for antidiabetic medications [37]. Third, our patients had a long duration of diabetes and were insulin users [32]. In general, the effects of education and lifestyle changes decrease with the duration of diabetes [38]. Fourth, there was no evidence of the most effective frequency of the intervention messages. We sent personalized intervention messages from HCPs every 2 weeks and automated general informative messages every other day. Patient satisfaction and accessibility are important for improving self-management efficiency, and the clinical course can be improved through personalized intervention [4]. More frequent, bidirectional, real-time communication with HCPs and patients would lead to more effective improvement in HbA_{1c} levels.

Although we did not observe remarkable improvement in HbA_{1c} levels over the long term, it is encouraging that the goal achievement rates for PA were higher in the intervention group at 26 weeks. When the target of 7500 steps per day was applied [39-41], the difference in goal achievement rates among the groups further increased (UC vs MC vs MPC: 25.9% vs 28.5% vs 30.2%). Given the lifelong management of T2DM, the small differences observed in the short term may increase in the future. Furthermore, in terms of the prevention of diabetic complications such as cardiovascular disease, PA cannot be overemphasized [40-42]. Finally, we expect that our study will provide more solid evidence of the short-term efficacy of mobile app–based diabetes management. In particular, in relation to the recent global public health crisis, the COVID-19 pandemic, this methodology is expected to contribute greatly in the future to promote the rapid introduction and diffusion of new digital health–related technologies such as telemedicine [43].

To maximize the effect of mHealth interventions, it is important to tailor the intervention in a patient-centered manner and evaluate user satisfaction [44]. Undoubtedly, more RCTs with longer follow-up periods should be conducted to evaluate the long-term effects of diabetes-related mobile apps and to confirm that the outcomes seen in initial studies are sustainable over time [22].

Conclusions

In summary, the use of iCareD apps for diabetes self-care can be considered an effective measure, especially when patients can communicate with HCPs [8]. Remote health data monitoring and real-time communication with patients supported self-care of diabetes, resulting in short-term improvement in HbA_{1c} levels. An mHealth system for patients with T2DM should be developed to support and motivate sustainable behavior changes in patients and to allow for an approach that is more tailored to individual needs.

Acknowledgments

This research was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute, funded by the Ministry of Health & Welfare, Republic of Korea (grant HI19C0784). The funders did not play any role in the study design, data collection and analysis, management preparation, or decision to publish.

Authors' Contributions

EYL contributed to the drafting and revision of the manuscript, supervision of the study, and acquisition of data. JSY and SAC contributed to the revision of the manuscript, supervision of the study, and acquisition of data. SYL and JHL contributed to the statistical methodology and data management. YBA and KHY contributed to the revision of the manuscript and supervision of the study. SHK contributed to the revision of the manuscript and design and supervision of the study. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Comparison of the intervention protocol in each group.
[DOCX File , 17 KB - jmir_v24i7e37430_app1.docx]

Multimedia Appendix 2
Hemoglobin A_{1c} (HbA_{1c}) level and HbA1c level changes from baseline to 12 and 26 weeks according to subgroup analysis.
[DOCX File , 21 KB - jmir_v24i7e37430_app2.docx]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2853 KB - jmir_v24i7e37430_app3.pdf]

References

1. Sun H, Saeedi P, Karuranga S, Pinkepank M, Ogurtsova K, Duncan BB, et al. IDF Diabetes Atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res Clin Pract 2022 Jan;183:109119. [doi: 10.1016/j.diabres.2021.109119] [Medline: 34879977]
2. Hur KY, Moon MK, Park JS, Kim SK, Lee SH, Yun JS, Committee of Clinical Practice Guidelines, Korean Diabetes Association. 2021 clinical practice guidelines for diabetes mellitus of the Korean Diabetes Association. Diabetes Metab J 2021 Jul;45(4):461-481 [FREE Full text] [doi: 10.4093/dmj.2021.0156] [Medline: 34352984]
3. American Diabetes Association Professional Practice Committee, Draznin B, Aroda VR, Bakris G, Benson G, Brown FM, et al. 9. Pharmacologic approaches to glycemic treatment: standards of medical care in diabetes-2022. Diabetes Care 2022 Jan 01;45(Suppl 1):S125-S143. [doi: 10.2337/dc22-S009] [Medline: 34964831]
4. Rhee SY, Kim C, Shin DW, Steinhubl SR. Present and future of digital health in diabetes and metabolic disease. Diabetes Metab J 2020 Dec;44(6):819-827 [FREE Full text] [doi: 10.4093/dmj.2020.0088] [Medline: 33389956]
5. Arora S, Peters AL, Burner E, Lam CN, Menchine M. Trial to examine text message-based mHealth in emergency department patients with diabetes (TExT-MED): a randomized controlled trial. Ann Emerg Med 2014 Jun;63(6):745-54.e6. [doi: 10.1016/j.annemergmed.2013.10.012] [Medline: 24225332]
6. Shan R, Sarkar S, Martin SS. Digital health technology and mobile devices for the management of diabetes mellitus: state of the art. Diabetologia 2019 Jun;62(6):877-887. [doi: 10.1007/s00125-019-4864-7] [Medline: 30963188]
7. Hyun MK, Lee JW, Ko SH, Hwang JS. Improving glycemic control in type 2 diabetes using mobile applications and e-coaching: a mixed treatment comparison network meta-analysis. J Diabetes Sci Technol (forthcoming) 2021 May 12:19322968211010153. [doi: 10.1177/19322968211010153] [Medline: 33980055]
8. Bonoto BC, de Araújo VE, Godói IP, de Lemos LL, 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 01;5(3):e4 [FREE Full text] [doi: 10.2196/mhealth.6309] [Medline: 28249834]
9. Orsama AL, Lähteenmäki J, Harno K, Kulju M, Wintergerst E, Schachner H, et al. Active assistance technology reduces glycosylated hemoglobin and weight in individuals with type 2 diabetes: results of a theory-based randomized trial. Diabetes Technol Ther 2013 Aug;15(8):662-669. [doi: 10.1089/dia.2013.0056] [Medline: 23844570]
10. 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]

11. Lee JH, Lim SY, Cha SA, Han CJ, Jung AR, Kim KR, et al. Short-term effects of the Internet-based Korea diabetes prevention study: 6-month results of a community-based randomized controlled trial. Diabetes Metab J 2021 Nov;45(6):960-965 [FREE Full text] [doi: 10.4093/dmj.2020.0225] [Medline: 33721974]
12. American Diabetes Association Professional Practice Committee, Draznin B, Aroda VR, Bakris G, Benson G, Brown FM, et al. 7. Diabetes technology: standards of medical care in diabetes-2022. Diabetes Care 2022 Jan 01;45(Suppl 1):S97-112. [doi: 10.2337/dc22-S007] [Medline: 34964871]
13. Lee EY, Yun JS, Cha SA, Lim SY, Lee JH, Ahn YB, et al. Personalized type 2 diabetes management using a mobile application integrated with electronic medical records: an ongoing randomized controlled trial. Int J Environ Res Public Health 2021 May 16;18(10):5300 [FREE Full text] [doi: 10.3390/ijerph18105300] [Medline: 34065775]
14. 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]
15. Hou C, Xu Q, Diao S, Hewitt J, Li J, Carter B. Mobile phone applications and self-management of diabetes: a systematic review with meta-analysis, meta-regression of 21 randomized trials and GRADE. Diabetes Obes Metab 2018 Aug;20(8):2009-2013. [doi: 10.1111/dom.13307] [Medline: 29582538]
16. Cho JH, Chang SA, Kwon HS, Choi YH, Ko SH, Moon SD, et al. Long-term effect of the Internet-based glucose monitoring system on HbA1c reduction and glucose stability: a 30-month follow-up study for diabetes management with a ubiquitous medical care system. Diabetes Care 2006 Dec;29(12):2625-2631. [doi: 10.2337/dc05-2371] [Medline: 17130195]
17. Platform Overview. Google Developers. URL: https://developers.google.com/fit/overview/ [accessed 2022-11-04]
18. Levy JC, Matthews DR, Hermans MP. Correct homeostasis model assessment (HOMA) evaluation uses the computer program. Diabetes Care 1998 Dec;21(12):2191-2192. [doi: 10.2337/diacare.21.12.2191] [Medline: 9839117]
19. HOMA2 Calculator. Diabetes Trials Unit, University of Oxford. 2019. URL: https://www.dtu.ox.ac.uk/homacalculator/ [accessed 2022-11-04]
20. Park SY, Gautier JF, Chon S. Assessment of insulin secretion and insulin resistance in human. Diabetes Metab J 2021 Sep;45(5):641-654 [FREE Full text] [doi: 10.4093/dmj.2021.0220] [Medline: 34610719]
21. 4 Digital Health App Trends to Consider for 2018. Liquid State. 2018 Jan 29. URL: https://liquid-state.com/digital-health-app-trends-2018/ [accessed 2022-11-30]
22. Doupis J, Festas G, Tsilivigos C, Efthymiou V, Kokkinos A. Smartphone-based technology in diabetes management. Diabetes Ther 2020 Mar;11(3):607-619 [FREE Full text] [doi: 10.1007/s13300-020-00768-3] [Medline: 31983028]
23. Block G, Azar KM, Romanelli RJ, Block TJ, Hopkins D, Carpenter HA, et al. Diabetes prevention and weight loss with a fully automated behavioral intervention by email, Web, and mobile phone: a randomized controlled trial among persons with prediabetes. J Med Internet Res 2015 Oct 23;17(10):e240 [FREE Full text] [doi: 10.2196/jmir.4897] [Medline: 26499966]
24. Harding JL, Pavkov ME, Magliano DJ, Shaw JE, Gregg EW. Global trends in diabetes complications: a review of current evidence. Diabetologia 2019 Jan;62(1):3-16. [doi: 10.1007/s00125-018-4711-2] [Medline: 30171279]
25. Waki K, Fujita H, Uchimura Y, Omae K, Aramaki E, Kato S, et al. DialBetics: a novel smartphone-based self-management support system for type 2 diabetes patients. J Diabetes Sci Technol 2014 Mar;8(2):209-215 [FREE Full text] [doi: 10.1177/1932296814526495] [Medline: 24876569]
26. Garabedian LF, Ross-Degnan D, Wharam JF. Mobile phone and smartphone technologies for diabetes care and self-management. Curr Diab Rep 2015 Dec;15(12):109 [FREE Full text] [doi: 10.1007/s11892-015-0680-8] [Medline: 26458380]
27. Wu Y, Yao X, Vespasiani G, Nicolucci A, Dong Y, Kwong J, et al. Mobile app-based interventions to support diabetes self-management: a systematic review of randomized controlled trials to identify functions associated with glycemic efficacy. JMIR Mhealth Uhealth 2017 Mar 14;5(3):e35 [FREE Full text] [doi: 10.2196/mhealth.6522] [Medline: 28292740]
28. Quinn CC, Clough SS, Minor JM, Lender D, Okafor MC, Gruber-Baldini A. WellDoc mobile diabetes management randomized controlled trial: change in clinical and behavioral outcomes and patient and physician satisfaction. Diabetes Technol Ther 2008 Jun;10(3):160-168. [doi: 10.1089/dia.2008.0283] [Medline: 18473689]
29. Agarwal P, Mukerji G, Desveaux L, Ivers NM, Bhattacharyya O, Hensel JM, et al. Mobile app for improved self-management of type 2 diabetes: multicenter pragmatic randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 10;7(1):e10321 [FREE Full text] [doi: 10.2196/10321] [Medline: 30632972]
30. Clements MA, Staggs VS. A mobile app for synchronizing glucometer data: impact on adherence and glycemic control among youths with type 1 diabetes in routine care. J Diabetes Sci Technol 2017 May;11(3):461-467 [FREE Full text] [doi: 10.1177/1932296817691302] [Medline: 28745097]
31. Tildesley HD, Mazanderani AB, Chan JH, Ross SA. Efficacy of A1C reduction using Internet intervention in patients with type 2 diabetes treated with insulin. Can J Diabetes 2011 Jan;35(3):250-253. [doi: 10.1016/s1499-2671(11)53007-2]
32. Chow N, Shearer D, Aydin Plaa J, Pottinger B, Pawlowska M, White A, et al. Blood glucose self-monitoring and internet diabetes management on A1C outcomes in patients with type 2 diabetes. BMJ Open Diabetes Res Care 2016 Apr 29;4(1):e000134 [FREE Full text] [doi: 10.1136/bmjdrc-2015-000134] [Medline: 27158516]

33. Davison BK, Quigg R, Skidmore PM. Pilot testing a photo-based food diary in nine- to twelve- year old- children from Dunedin, New Zealand. Nutrients 2018 Feb 20;10(2):240 [FREE Full text] [doi: 10.3390/nu10020240] [Medline: 29461499]
34. Fuller NR, Fong M, Gerofi J, Ferkh F, Leung C, Leung L, et al. Comparison of an electronic versus traditional food diary for assessing dietary intake-a validation study. Obes Res Clin Pract 2017;11(6):647-654. [doi: 10.1016/j.orcp.2017.04.001] [Medline: 28420588]
35. Kumar RB, Goren ND, Stark DE, Wall DP, Longhurst CA. Automated integration of continuous glucose monitor data in the electronic health record using consumer technology. J Am Med Inform Assoc 2016 May;23(3):532-537 [FREE Full text] [doi: 10.1093/jamia/ocv206] [Medline: 27018263]
36. Holzinger A, Searle G, Nischelwitzer A. On some aspects of improving mobile applications for the elderly. In: Proceedings of the 4th International Conference on Universal Access in Human-Computer Interaction. 2007 Presented at: UAHCI '07; July 22-27, 2007; Beijing, China p. 923-932. [doi: 10.1007/978-3-540-73279-2_103]
37. Holmen H, Torbjørnsen A, Wahl AK, Jenum AK, Småstuen MC, Arsand E, et al. A mobile health intervention for self-management and lifestyle change for persons with type 2 diabetes, part 2: one-year results from the Norwegian randomized controlled trial RENEWING HEALTH. JMIR Mhealth Uhealth 2014 Dec 11;2(4):e57 [FREE Full text] [doi: 10.2196/mhealth.3882] [Medline: 25499872]
38. Ko SH, Park SA, Cho JH, Ko SH, Shin KM, Lee SH, et al. Influence of the duration of diabetes on the outcome of a diabetes self-management education program. Diabetes Metab J 2012 Jun;36(3):222-229 [FREE Full text] [doi: 10.4093/dmj.2012.36.3.222] [Medline: 22737662]
39. Lee IM, 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 Aug 01;179(8):1105-1112 [FREE Full text] [doi: 10.1001/jamainternmed.2019.0899] [Medline: 31141585]
40. Hajna S, Ross NA, Dasgupta K. Steps, moderate-to-vigorous physical activity, and cardiometabolic profiles. Prev Med 2018 Feb;107:69-74 [FREE Full text] [doi: 10.1016/j.ypmed.2017.11.007] [Medline: 29126915]
41. Paluch AE, Bajpai S, Bassett DR, Carnethon MR, Ekelund U, Evenson KR, Steps for Health Collaborative. Daily steps and all-cause mortality: a meta-analysis of 15 international cohorts. Lancet Public Health 2022 Mar;7(3):e219-e228 [FREE Full text] [doi: 10.1016/S2468-2667(21)00302-9] [Medline: 35247352]
42. Saint-Maurice PF, Troiano RP, Bassett Jr DR, Graubard BI, Carlson SA, Shiroma EJ, et al. Association of daily step count and step intensity with mortality among US adults. JAMA 2020 Mar 24;323(12):1151-1160 [FREE Full text] [doi: 10.1001/jama.2020.1382] [Medline: 32207799]
43. Keesara S, Jonas A, Schulman K. Covid-19 and health care's digital revolution. N Engl J Med 2020 Jun 04;382(23):e82. [doi: 10.1056/NEJMp2005835] [Medline: 32240581]
44. Cahn A, Akirov A, Raz I. Digital health technology and diabetes management. J Diabetes 2018 Jan;10(1):10-17. [doi: 10.1111/1753-0407.12606] [Medline: 28872765]

Abbreviations

EMR: electronic medical record
HbA_{1c}: hemoglobin A_{1c}
HCP: health care provider
ITT: intention-to-treat
MC: mobile diabetes self-care
mHealth: mobile health
MPC: mobile diabetes self-care with personalized, bidirectional feedback from physicians
PA: physical activity
RCT: randomized controlled trial
SMBG: self-monitoring of blood glucose
T2DM: type 2 diabetes mellitus
UC: usual care

Edited by R Kukafka; submitted 21.02.22; peer-reviewed by Q Xu, N Maglaveras; comments to author 05.04.22; revised version received 26.05.22; accepted 08.06.22; published 28.07.22.

Please cite as:
Lee EY, Cha SA, Yun JS, Lim SY, Lee JH, Ahn YB, Yoon KH, Hyun MK, Ko SH
Efficacy of Personalized Diabetes Self-care Using an Electronic Medical Record–Integrated Mobile App in Patients With Type 2 Diabetes: 6-Month Randomized Controlled Trial
J Med Internet Res 2022;24(7):e37430
URL: https://www.jmir.org/2022/7/e37430
doi:10.2196/37430
PMID:35900817

Original Paper

Improving Health Knowledge Through Provision of Free Digital Health Education to Rural Communities in Iringa, Tanzania: Nonrandomized Intervention Study

Christine Holst[1,2], MPhil; Dominik Stelzle[3], MPH, MD, Dr med; Lien My Diep[4], MSc; Felix Sukums[5], PhD; Bernard Ngowi[6,7], MD, PhD; Josef Noll[8,9], PhD; Andrea Sylvia Winkler[1,3], MD, Dr med, PhD

[1]Centre for Global Health, Department of Community Medicine and Global Health, Institute of Health and Society, University of Oslo, Oslo, Norway

[2]Centre for Epidemic Interventions Research, Norwegian Institute of Public Health, Oslo, Norway

[3]Center for Global Health, Department of Neurology, Technical University of Munich, Munich, Germany

[4]Oslo Centre for Biostatistics & Epidemiology, Oslo University Hospital, Oslo, Norway

[5]Directorate of Information and Communication Technology, Muhimbili University of Health and Allied Sciences, Dar es Salaam, United Republic of Tanzania

[6]Muhimbili Medical Research Centre, National Institute for Medical Research, Dar es Salaam, United Republic of Tanzania

[7]Mbeya College of Health and Allied Sciences, University of Dar es Salaam, Mbeya, United Republic of Tanzania

[8]Basic Internet Foundation, Kjeller, Norway

[9]Department of Technology Systems, University of Oslo, Oslo, Norway

Corresponding Author:
Christine Holst, MPhil
Centre for Global Health
Department of Community Medicine and Global Health, Institute of Health and Society
University of Oslo
PO Box 1130
Blinderen
Oslo, 0318
Norway
Phone: 47 48234044
Email: christine.holst@medisin.uio.no

Abstract

Background: Community health education is one of the most effective measures to increase health literacy worldwide and can contribute to the achievement of specific targets of the Sustainable Development Goal 3. Digitalized health education materials can improve health knowledge as a dimension of health literacy and play an important role in disease prevention in rural sub-Saharan settings.

Objective: The objective of this research is to assess the effect of a digital health education intervention on the uptake and retention of knowledge related to HIV/AIDS, tuberculosis (TB), and *Taenia solium* (neuro)cysticercosis and taeniosis in rural communities in Iringa, Tanzania.

Methods: We conducted a nonrandomized intervention study of participants aged 15 to 45 years, randomly selected from 4 villages in Iringa, Tanzania. The intervention consisted of 2 parts. After the baseline assessment, we showed the participants 3 animated health videos on a tablet computer. After a period of 6 months, free access to community information spots (InfoSpots) with an integrated digital health education platform was provided to the intervention villages. Participants in the control group did not receive the intervention. The primary outcome was the difference in disease knowledge between the intervention and control groups, 12 months after baseline. Data were collected using an open-ended questionnaire, with correct or incorrect answers before and after intervention.

Results: Between April and May 2019, a total of 600 participants were recruited into the intervention (n=298, 49.7%) or control (n=302, 50.3%) groups. At baseline, no statistically significant differences in knowledge of the target diseases were observed. At 12 months after intervention, knowledge about HIV/AIDS, TB, and *T. solium* (neuro)cysticercosis and taeniosis was 10.2% (95% CI 5.0%-15.4%), 12% (95% CI 7.7%-16.2%), and 31.5% (95% CI 26.8%-36.2%) higher in the intervention group than in the control group, respectively. In all 4 domains (transmission, symptoms, treatment, and prevention), an increase in knowledge

was observed in all the 3 diseases, albeit to varying degrees. The results were adjusted for potential confounders, and the significance of the primary results was maintained in the sensitivity analysis to assess dropouts. The participants who reported using the InfoSpots in the 12-month assessment further increased their knowledge about the target diseases by 6.8% (HIV/AIDS), 7.5% (TB), and 13.9% higher mean proportion of correct answers compared with the participants who did not use the InfoSpots.

Conclusions: Digital health education based on animated health videos and the use of free InfoSpots has significant potential to improve health knowledge, especially in rural areas of low- and middle-income countries.

Trial Registration: ClinicalTrials.gov NCT03808597; https://clinicaltrials.gov/ct2/show/NCT03808597

International Registered Report Identifier (IRRID): RR2-10.2196/25128

(J Med Internet Res 2022;24(7):e37666) doi:10.2196/37666

KEYWORDS

digital health; digital health promotion; eHealth; mobile health; mHealth; Tanzania; health education; HIV/AIDS; tuberculosis; cysticercosis; tapeworm; mobile phone

Introduction

Background

Health education, defined as "any combination of learning experiences designed to facilitate voluntary adaptations of behaviour conducive to health" [1], encourages behavior that promotes good health and prevents disease [2]. Health literacy is a concept defined differently across the literature but can be summarized as "an asset for improving people's empowerment within the domains of health care, disease prevention and health promotion" [3]. Health literacy acquired through health education affects individual, family, and community health and has great potential for contributing to the achievement of specific targets of the Sustainable Development Goal (SDG; SDG 3) [4]. Health knowledge can be seen as a consequence, antecedent, or dimension of health literacy [5] and includes specific information such as prevalence, risk factors, transmission, and prevention [6]. The vast majority of people in rural communities in sub-Saharan Africa, Tanzania being the focus of this study, need health education, which is primarily provided by community health workers and individuals at health facilities, schools, and nongovernmental organizations.

The southern and eastern regions of Africa are most affected by HIV/AIDS. In 2020, 20.6 million people were living with HIV/AIDS, and 670,000 new infections and 310,000 AIDS-related deaths were registered [7]. Tuberculosis (TB) claimed 1.5 million lives worldwide in 2020 [8]. In addition, TB is the leading cause of death among HIV-infected individuals, often owing to late diagnosis, nonadherence to medication, and drug resistance. *Taenia solium* cysticercosis and taeniosis (TSCT) is considered a neglected tropical disease and is a food-borne parasitic disease that not only has a significant impact on human health by causing neurological signs or symptoms, including epileptic seizures and intestinal infection resulting in anemia and malnutrition, but also has an impact on animal health and community livelihood and can therefore be considered a one health disease par excellence [9].

The digital health landscape in Tanzania is changing in parallel with increasing mobile phone use and internet penetration. In December 2021, the country had 54.1 million mobile phone subscribers [10], with most connections in rural areas based on 2G. Coupled with significant economic growth and Tanzania's recent shift from low-income to lower-middle–income status [11], there is great potential to reach people through their mobile phones. Adapting smart devices for health education in resource-poor areas is promising and has proven effective [12].

Few studies have addressed the development [13] and use of animated health videos (animations) to improve client knowledge in this region [14-18]. Investigating the impact of a digital health education platform in rural sub-Saharan areas with low internet connectivity is novel and forms the basis of our study.

Objectives

The objective of the study in "the nondiscriminating access for digital inclusion" project ("DigI study") was to assess the uptake of knowledge with regard to HIV/AIDS, TB, and TSCT and its retention over time through a digital health education intervention with animations and provision of free access to community information spots (InfoSpots) with an integrated digital health education platform.

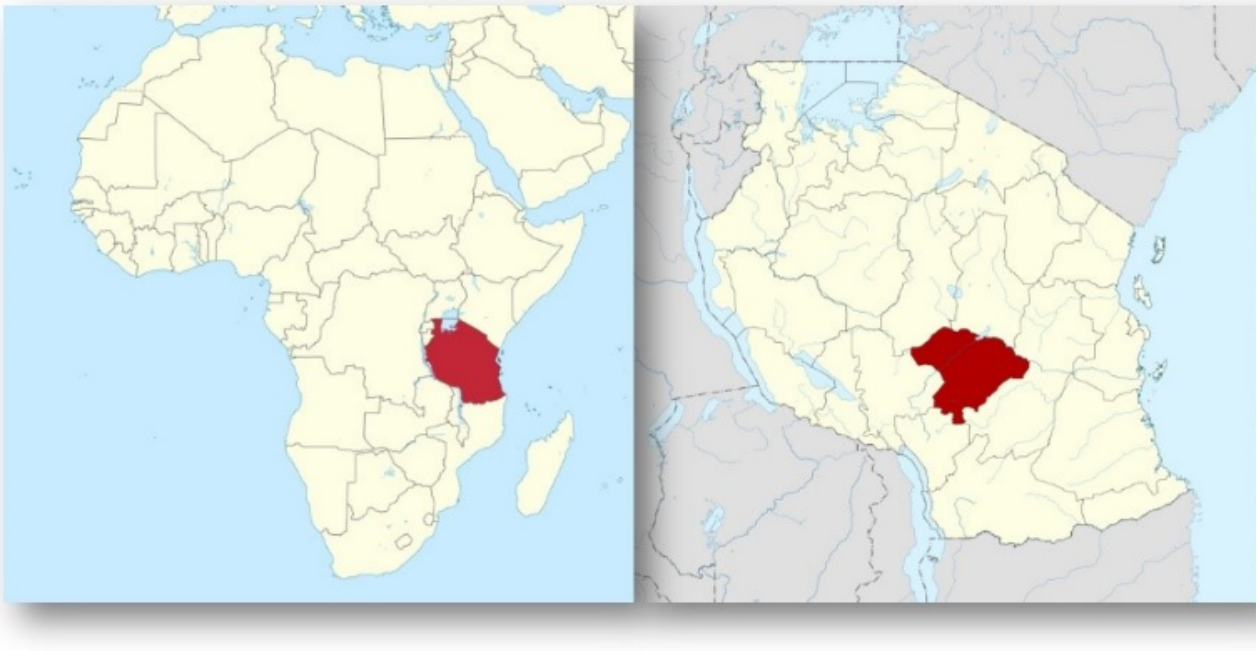
Methods

Study Design and Setting

As published in the DigI study protocol [19], this was a nonrandomized intervention study formatted as a community-based and longitudinal study. The study was undertaken in 4 villages with their 21 subvillages in the Iringa region located in the southern highlands of mainland Tanzania. The Iringa region is endemic to HIV/AIDS [20], TB [21], and TSCT [22], making prevention strategies for the target diseases essential. The Iringa district is a rural area in the northern Iringa region, where agriculture and livestock are the main sources of income [23]. All the 4 villages included in this study had health facilities.

Households in Migoli and Izazi (Figure 1) became the intervention group, whereas households in Kimande and Idodi formed the control group. The intervention and control villages were purposively chosen based on technical installation reasons and a traveling distance of >2.5 hours by car between the intervention and control villages.

Figure 1. Geographic location of the study site. From left to right: Tanzania within Africa and the Iringa district within Tanzania. Both illustrations are from Wikimedia Commons contributors (CC BY-SA 3.0).



Participants

In April and May 2019, a total of 991 households were assessed for their eligibility to participate in the study. People aged between 15 and 45 years living permanently in randomly selected households from all subvillages were eligible for inclusion in the study. People who planned to move to areas outside the villages during the data collection period were not eligible to participate. A female or male member per household was randomly selected using the Kish-grid method [24] to participate in the study based on multilevel sampling.

Intervention

The DigI study intervention consists of 2 parts: animations with key messages on HIV/AIDS, TB, and TSCT and the digital health education platform. The first part of the intervention—3 stories told with animations about the target diseases (Figure 2)—was presented to participants by Tanzanian members of the study team. The animations were shown on a tablet to the participants privately in their homes after the baseline assessment using the questionnaire. The animations, lasting between 3 and 7 minutes, included key messages for 4 domains (transmission, symptoms, treatment, and prevention) in each disease (HIV/AIDS, TB, and TSCT).

The DigI team applied an interdisciplinary approach when creating digital health messages based on a process involving local stakeholders and government-approved health promotion materials in a nondigital format [25]. All the animations were bilingual in Swahili and English [26]. All participants in the intervention group viewed animations in the Tanzanian Swahili language.

The second, unsupervised and voluntary part of the intervention started 6 months after the baseline assessment, with the rollout of the digital health education platform (the platform) in the already established InfoSpots in the intervention villages. The platform [27] could be accessed openly and free of charge via the participants' own smartphones or via public tablets provided at the InfoSpots. In addition to the animations in the first part of the intervention, the platform contained additional text information, graphics, and quizzes. We installed 5 InfoSpots in Migoli and Izazi, covering village offices and health facilities as well as Nyerere High School in Migoli. The animations (English versions) and a screenshot of the platform can be found in Figure 3 and Multimedia Appendices 1-3.

The control villages did not receive any intervention. Communities in the region usually access general information, health information, and education and communication materials through various channels. These channels include radio, television, newspapers, community meetings, and outreach from schools via children as well as when visiting health facilities.

Figure 2. Examples of illustrations of key messages from the animations. Important key messages were, for example, “never share sharp objects” and “avoid unsafe blood transfusions” for HIV/AIDS, “cover mouth and nose with paper or cloth when coughing or sneezing” for tuberculosis, and “cook meat well” for Taenia solium cysticercosis and taeniosis. These key messages were well illustrated, narrated, and repeated in the animations.

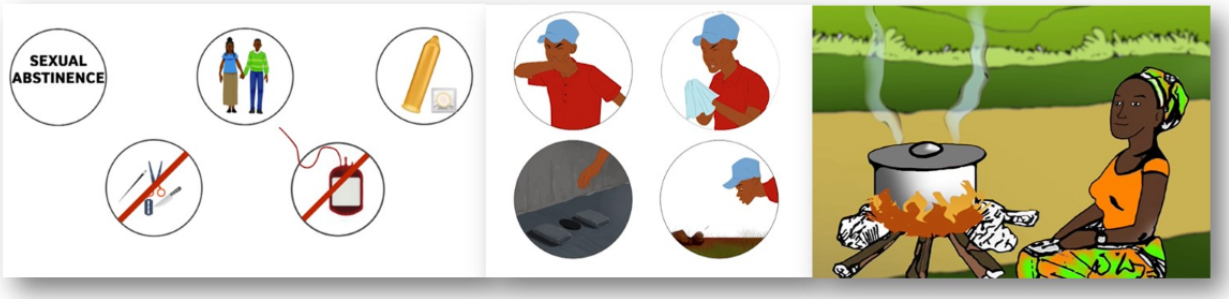
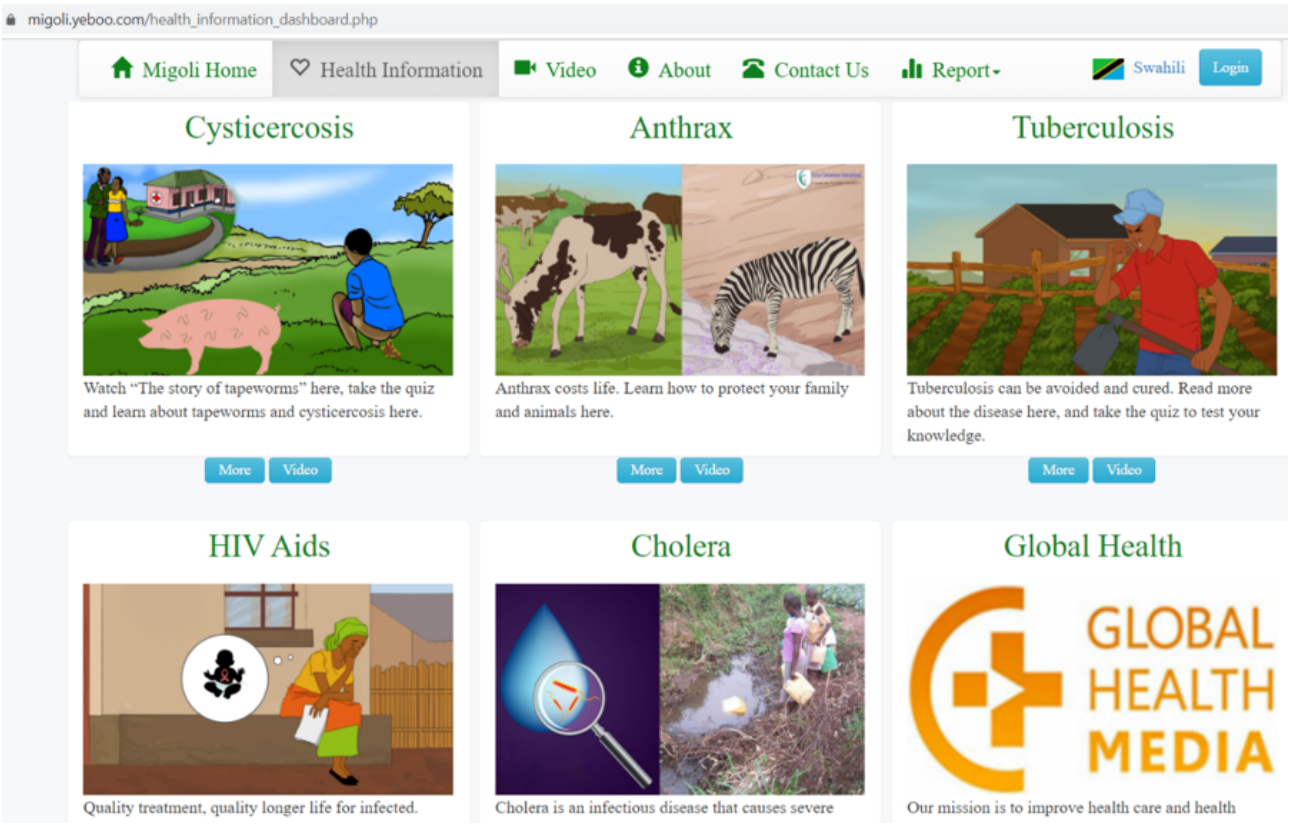


Figure 3. Screenshot of the platform.



Questionnaire

An open-ended Swahili-based questionnaire was used for data collection. The questionnaire was designed by Tanzanian and international experts in English (Multimedia Appendix 4), based on questions from previous studies [28,29] (The International Organisation for Migration, unpublished data, January 2014) and expanded through additional questions derived from approved health information provided by the Tanzanian Ministry of Health. The questions were translated to Swahili and back-translated into English again. A pilot study (in Swahili) was conducted with 50 participants before data collection [19]. A total of 7 questions were adapted to increase the participants’ comprehensibility. The questionnaire contained questions that could have more than one correct answer. Answers to the questions were binary (correct or incorrect. The questions were asked face to face and the enumerator entered the participant’s answers into a digital form. To analyze the health knowledge of the participants, we used 12 HIV/AIDS-knowledge questions, with 23 possible correct answers, 8 TB questions with 30 possible correct answers, and 13 TSCT questions with 22 possible correct answers. The overall Cronbach α, per disease, ranged between .79 and .87.

Outcomes

The primary outcome of the DigI study was the difference in the overall proportion of correct answers per target disease between the intervention and control groups 12 months after baseline assessment. The main hypothesis was that the proportion of correct answers in the intervention group would differ from the proportion of correct answers in the control group at 3 and 12 months. The secondary outcome was the difference in the proportion of correct answers in the 4 domains of each disease: transmission, symptoms, treatment, and

prevention. The secondary hypothesis was that there would be differences among the groups' proportion of correct answers per domain. For all outcome variables, the questionnaire and corresponding correct responses were used.

To investigate the effect of the digital health education intervention, we assessed the disease knowledge of both the intervention and control groups at baseline and at 3 and 12 months after the first part of the intervention. Participants in the intervention group were further asked questions related to their views regarding animations at the 3-month follow-up and their use of the InfoSpots at the 12-month follow-up.

Statistical Methods

For the sample size calculation, we assumed a proportion of 50% of correct answers in the group without intervention and a difference among the groups of 15% to 20% after the intervention. The power was set to 80% and a 2-sided significance level of 5% (using Bonferroni correction) was applied. The target sample size was 460. To increase the precision of the key estimates and account for dropouts, 600 participants were recruited.

Sociodemographic sample characteristics were described by frequency with percentage or mean with SD. Differences in sociodemographic characteristics among the groups were tested using Pearson chi-square test (for categorical variables) and a 2-independent sample 2-tailed *t* test (for age). For each disease, the total number of correct answers was summed for each participant and divided by the maximum number of correct answers to generate the overall proportion of correct answers. The same calculations were performed to generate the proportion of correct answers for each domain per disease. Assuming missing data at random, a mixed effects linear regression model for continuous outcomes was used to examine the difference in the average proportion of correct answers between the intervention and control groups. The analysis was performed with and without adjusting for age, sex, education, and occupation as potential confounders that could have influenced the outcome. In the mixed model of linear regression, the subvillage, household, and time point were specified as random variables and the group, as a fixed variable. Unstructured and exchangeable variance-covariance structures for the households and time were chosen. To assess the impact of time with and without adjusting, an interaction term "group×time" was added to the mixed model. The results of the analysis using mixed effects linear regression were reported as the average difference among the groups, and 95% CI were calculated. The *P* value was also reported for additional information. Loss to follow-up was addressed by sensitivity analysis assuming a worst-case scenario, with all answers being incorrect after the dropout of the participant. The differences in the proportion of correct answers between the participants who used the InfoSpots and those who did not were described using mean and SD. Analyses were performed using Stata/SE (version 16; StataCorp) [30].

Ethics Approval

The Tanzanian National Health Research Ethics Sub-Committee granted ethical clearance (NIMR/HQ/R.8a/Vol. IX/2947) in November 2018 and extended its clearance in February 2020. The Tanzania Commission for Science and Technology issued a research permit in March 2019. The DigI study was registered at ClinicalTrials.gov (NCT03808597) before enrolling the first participant. Each participant was informed orally and in writing in Swahili. Informed consent to participate in the study was obtained in writing, and parental consent was required for participants aged <18 years.

Results

Study Population

A total of 600 participants were recruited into the study, of whom 560 (93.3%) were interviewed at 3-month follow-up and 493 (82.2%) at 12-month follow-up (see Figure 4 for a flowchart of the study).

The descriptions of the 2 study populations are presented in Table 1. The main differences were in education and occupation, which we adjusted for in our analysis.

Figure 4. Flowchart of the study. The figure shows the number of households assessed for inclusion before a random generator selected the 298 households for the intervention group and 302 households for the control group. One participant per household was included in the study. Further, the figure shows the remaining participants at the 3- and 12-month assessment.

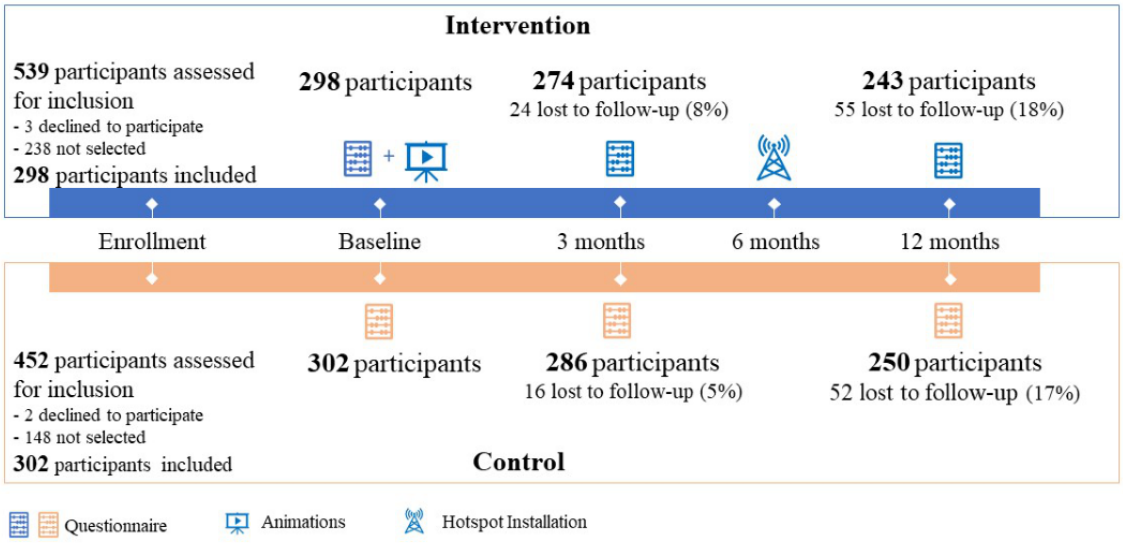


Table 1. Baseline demographic characteristics of participants aged 15 to 45 years living in the intervention and control villages.

Characteristics	Total (N=600)	Intervention (n=298)	Control (n=302)
Sex, n (%)			
Female	362 (60.3)	193 (64.8)	169 (56)
Male	238 (39.7)	105 (35.2)	133 (44)
Age (years), mean (SD)	30.0 (8.3)	28.7 (8.1)	30.8 (8.3)
Village, n (%)			
Migoli	173 (28.8)	173 (58.1)	N/A[a]
Izazi	125 (20.8)	125 (41.9)	N/A
Idodi	150 (25)	N/A[a]	150 (49.7)
Kimande	152 (25.3)	N/A	152 (50.3)
Education, n (%)			
No education	79 (13.2)	56 (18.8)	23 (7.6)
Primary School	427 (71.2)	181 (60.7)	246 (81.5)
Secondary or higher education	94 (15.7)	61 (20.4)	33 (11)
Occupation, n (%)			
Petty traders	151 (25.2)	135 (45.3)	16 (5.3)
Farmers	321 (53.5)	65 (21.8)	256 (84.8)
Other	128 (21.3)	98 (32.9)	17 (9.9)

[a]N/A: not applicable.

Knowledge Outcomes

At baseline, all (600/600, 100%) participants reported that they had heard of HIV/AIDS. Overall, 91% (546/600) of the participants had heard of TB, but only 9% (54/600) of the participants had heard of cysticercosis, and 13.2% (79/600) of the participants had heard of the pork tapeworm. Overall, only 3% (18/600) of the participants said that they knew what pork with cysticercosis looked like. The baseline knowledge, calculated as the unadjusted mean percentage of correct answers, was similar in both groups, although it was slightly higher for TB and TSCT in the intervention group than in the control group. The overall proportions of correct answers at the disease level are presented in Table 1 and Figure 5. The equivalent responses broken down at the domain level are presented in Table 2.

At the 3-month assessment, correct answers regarding HIV/AIDS in the intervention group increased to a mean

percentage of 54.7% (SD 17%), whereas they remained at 47.1% (SD 16.6%) in the control group (adjusted difference 8.3, 95% CI 3.2-13.4). The intervention effect was larger in the TB section, with an increase from baseline of 15.5% to 34.4% (SD 15.4%) in the intervention group compared with a 1.5% increase in the control group (adjusted difference 15.8, 95% CI 11.7-20.0). For TSCT, the first part of the intervention led to a 31.3% knowledge increase, reaching 41% (SD 20.2%) in the intervention group, whereas only a limited increase of 2.5% was measured in the control group (adjusted difference 31.2, 95% CI 26.9-35.6) after 3 months.

At the 3-month assessment, 97.1% (266/274) of the participants in the intervention group reported learning through animations, and 74.1% (203/274) reported having discussed health messages with friends and family.

Figure 5. Box plot and density plots for knowledge and by disease, group, and time point: (A) HIV/AIDS, (B) tuberculosis, (C) Taenia solium cysticercosis and taeniosis.

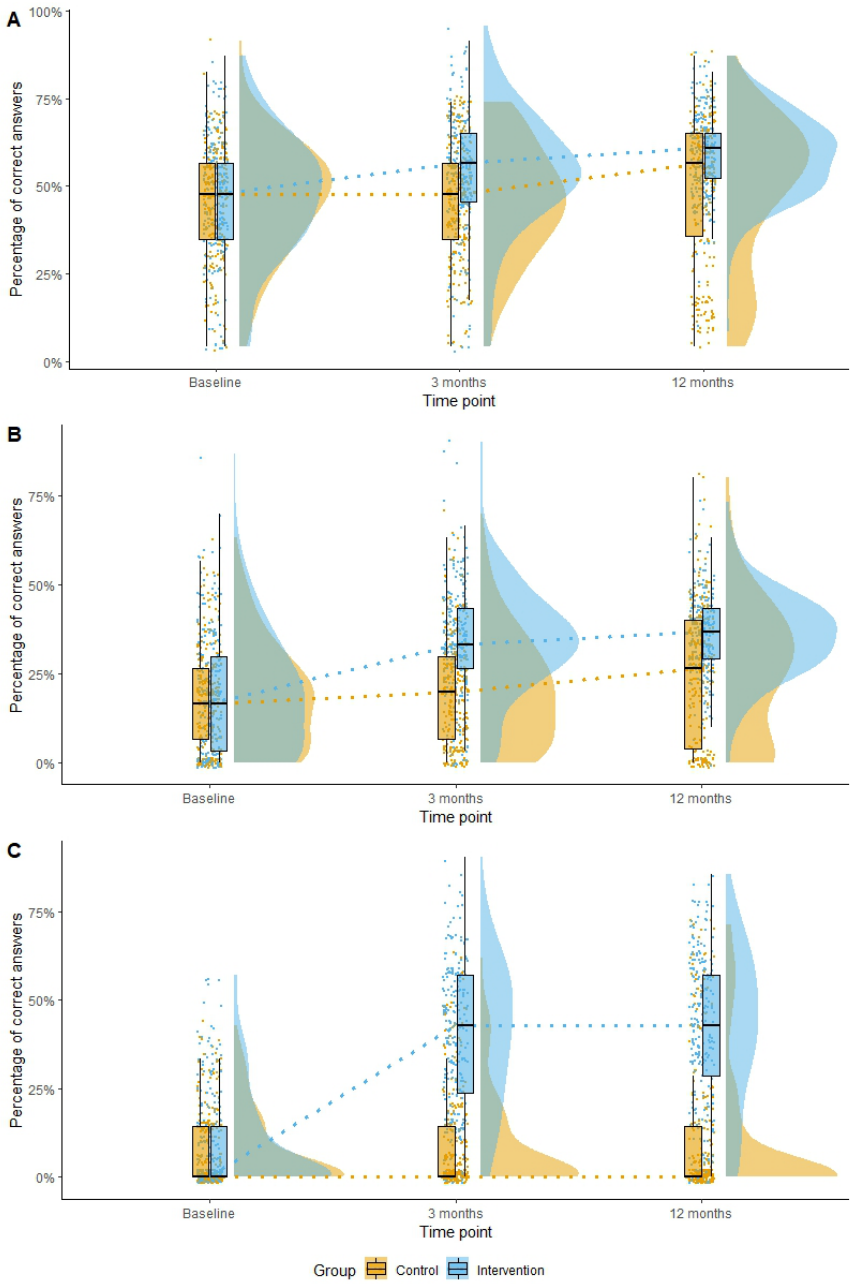


Table 2. Overall intervention effect for knowledge about the target diseases and mean percentage differences among the groups (N=600).

Target disease and time point	Correct answers, mean (SD)		Crude difference, mean (95% CI)	Adjusted difference, mean (95% CI)	*P* value (adjusted model)
	Intervention	Control			
Overall HIV/AIDS					<.001
Baseline	47.6 (17.4)	48.3 (15.6)	−0.1 (−6.5 to 6.3)	0.3 (−4.7 to 5.4)	
3 months	54.7 (17.0)	47.1 (16.6)	8.0 (1.6 to 14.4)	8.3 (3.2 to 13.4)	
12 months	59.8 (11.9)	50.2 (21.1)	9.9 (3.4 to 16.3)	10.2 (5.0 to 15.4)	
Overall tuberculosis					<.001
Baseline	18.9 (16.2)	18.3 (14.7)	1.3 (−3.6 to 6.1)	2.0 (−2.1 to 6.1)	
3 months	34.4 (15.4)	19.8 (15.0)	15.2 (10.3 to 20.0)	15.8 (11.7 to 20.0)	
12 months	35.8 (11.5)	25.1 (18.5)	11.3 (6.3 to 16.2)	12.0 (7.7 to 16.2)	
Overall *Taenia solium* **cysticercosis and taeniosis**					<.001
Baseline	9.7 (14.9)	7.6 (12.1)	2.4 (−1.7 to 6.4)	2.6 (−1.6 to 6.8)	
3 months	41.0 (20.2)	10.1 (15.1)	31.1 (26.9 to 35.3)	31.2 (26.9 to 35.6)	
12 months	42.1 (22.0)	11.0 (20.0)	31.4 (26.8 to 35.9)	31.5 (26.8 to 36.2)	

Participants in the intervention group were able to maintain their knowledge of the disease, as documented at the 12-month assessment. Participants in the control group were found to have a slight increase in knowledge of 1.9% (HIV/AIDS), 6.8% (TB), and 3.4% (TSCT) from baseline, which was significantly lower than the increase of 12.2% for HIV/AIDS (adjusted difference 10.2, 95% CI 5.0-15.4), 16.9% for TB (adjusted difference 12.0, 95% CI 7.7-16.2), and 32.4% for TSCT (adjusted difference 31.5, 95% CI 26.8-36.2) in the intervention group. After 12 months, participants in the intervention group retained a mean proportion of correct answers of 59.8% for HIV/AIDS, 35.8% for TB, and 42.1% for TSCT.

The intervention effect was statistically significant in the crude analyses and remained statistically significant after adjusting for age, sex, education, and occupation. Figure 5 shows the corresponding raincloud plot illustrating the median percentage of correct answers in both groups at all time points.

The intervention effect in the transmission HIV/AIDS domain was statistically significant, with an adjusted mean difference of 6% (95% CI 1.7%-10.4%) after 3 months and 6.4% (95% CI 1.9%-10.9%) after 12 months. In this domain, the effect remained unchanged over time (*P*=.06), in contrast to the other domains, where the effect either increased or decreased between 3 and 12 months. The effect was significant for all TB domains, although knowledge increase at 12 months was reduced compared with that at 3 months. However, as with the overall scores, the strongest effect was observed for TSCT, for which there were low baseline levels of disease knowledge. The adjusted mean percentage difference between the 2 groups reached 40% in 3 out of the 4 domains after 3 months and in 2 out of the 4 domains after 12 months. See Table 3 for all domain level values.

A sensitivity analysis assessing the dropout of 107 participants (n=55, 51% in the intervention group and n=52, 49% in the control group) and assuming a worst-case scenario showed an unchanged effect for HIV/AIDS, a 3% reduction in the intervention effect for TB, and a 5.8% reduction in the intervention effect for TSCT.

At the 12-month assessment, 30% (73/243) of the participants in the intervention group reported having accessed the InfoSpots with the health education platform. Of these, 53% (39/73) had viewed HIV/AIDS animation, 37% (27/73) had TB animation, and 44% (32/73) had TSCT animation. Those who visited the InfoSpots further increased their knowledge by 8.5% (HIV/AIDS), 6.3% (TB), and 12.1% (TSCT), whereas participants who did not visit the InfoSpots remained at the same knowledge level (Table 4).

Table 3. Intervention effect and mean percentage differences among the groups at the domain level.

Domains and time point	Correct answers, mean (SD)		Crude difference, mean (95% CI)	Adjusted difference, mean (95% CI)	*P* value (adjusted model)
	Intervention	Control			
HIV/AIDS					
Transmission					.06
Baseline	49.5 (20.0)	47.8 (18.7)	2.0 (−4.0 to 7.9)	1.9 (−2.3 to 6.2)	
3 months	54.0 (19.4)	47.8 (20.2)	6.2 (0.2 to 12.1)	6.0 (1.7 to 10.4)	
12 months	59.2 (16.8)	52.5 (21.3)	6.4 (0.4 to 12.5)	6.4 (1.9 to 10.9)	
Symptoms					<.001
Baseline	79.2 (33.4)	81.1 (30.9)	−1.4 (−11.5 to 8.7)	−2.2 (−11.1 to 6.8)	
3 months	88.0 (28.3)	71.8 (37.1)	16.5 (6.4 to 26.7)	15.5 (6.5 to 24.6)	
12 months	93.6 (19.0)	75.0 (40.6)	18.8 (8.5 to 29.1)	17.9 (8.6 to 27.1)	
Treatment					<.001
Baseline	70.5 (29.5)	73.7 (27.6)	−2.4 (−12.6 to 7.8)	−2.8 (−11.9 to 6.3)	
3 months	77.2 (25.1)	67.3 (30.5)	10.6 (0.4 to 20.8)	10.1 (1.0 to 19.2)	
12 months	90.7 (18.1)	72.8 (37.2)	18.5 (8.1 to 28.8)	18.1 (8.8 to 27.3)	
Prevention					<.001
Baseline	35.3 (19.2)	36.8 (17.7)	−0.8 (−6.6 to 5.1)	0.2 (−4.7 to 5.1)	
3 months	43.9 (19.1)	37.5 (16.3)	7.0 (1.2 to 12.8)	7.9 (2.9 to 12.8)	
12 months	46.8 (15.6)	39.0 (19.9)	8.4 (2.5 to 14.2)	9.2 (4.2 to 14.3)	
Tuberculosis					
Transmission					<.001
Baseline	17.6 (22.9)	17.0 (17.3)	1.6 (−2.9 to 6.1)	2.1 (−1.7 to 6.0)	
3 months	29.6 (17.5)	18.8 (14.5)	11.5 (7.3 to 15.7)	12.1 (8.5 to 15.6)	
12 months	28.7 (14)	21.1 (16.2)	8.5 (4.2 to 12.7)	9.0 (5.3 to 12.6)	
Symptoms					<.001
Baseline	14.4 (15.5)	13.8 (13.4)	1.3 (−3.2 to 5.7)	1.8 (−2.1 to 5.6)	
3 months	28.7 (18.1)	15.2 (14.5)	14.0 (9.5 to 18.5)	14.4 (10.5 to 18.4)	
12 months	28.1 (15.0)	19.9 (18.4)	8.6 (4.0 to 13.3)	9.1 (5.0 to 13.2)	
Treatment					<.001
Baseline	54.8 (38.3)	53.8 (37.4)	2.3 (−9.0 to 13.5)	4.5 (−6.1 to 15.0)	
3 months	82.2 (28.6)	53.6 (38.4)	29.5 (18.5 to 40.5)	31.6 (21.2 to 41.9)	
12 months	89.0 (20.7)	61.6 (40.7)	28.4 (17.2 to 39.6)	30.4 (19.9 to 41.0)	
Prevention					<.001
Baseline	13.6 (18.5)	13.5 (16.9)	0.5 (−4.1 to 5.1)	1.1 (−3.0 to 5.2)	
3 months	29.2 (16.3)	15.3 (16.2)	14.1 (9.6 to 18.7)	14.6 (10.5 to 18.7)	
12 months	33.2 (13.4)	22.5 (19.8)	10.9 (6.3 to 15.6)	11.4 (7.3 to 15.6)	
Taenia solium **cysticercosis and taeniosis**					
Transmission					<.001
Baseline	8.7 (19.4)	4.7 (14.1)	4.2 (−1.3 to 9.7)	4.1 (−1.5 to 9.7)	
3 months	50.1 (28.7)	8.3 (19.1)	42.0 (36.2 to 47.7)	41.7 (35.9 to 47.5)	
12 months	51.4 (27.0)	10.0 (23.4)	41.5 (35.4 to 47.6)	41.2 (35.0 to 47.3)	
Symptoms					<.001

Domains and time point	Correct answers, mean (SD)		Crude difference, mean (95% CI)	Adjusted difference, mean (95% CI)	*P* value (adjusted model)
	Intervention	Control			
Baseline	6.0 (9.9)	5.6 (8.8)	0.9 (−1.8 to 3.7)	1.0 (−1.7 to 3.8)	
3 months	22.5 (19.9)	7.6 (10.6)	15.3 (12.4 to 18.3)	15.4 (12.4 to 18.5)	
12 months	27.1 (22.2)	8.1 (15.4)	19.5 (16.2 to 22.8)	19.6 (16.1 to 23.1)	
Treatment					<.001
Baseline	7.4 (26.2)	5.0 (21.8)	2.4 (−3.3 to 8.0)	2.8 (−3.5 to 9.2)	
3 months	65.6 (47.6)	16.2 (36.9)	49.3 (43.2 to 55.7)	49.5 (42.5 to 56.5)	
12 months	72.8 (44.6)	26.0 (44.0)	46.5 (39.0 to 53.9)	46.7 (38.7 to 54.6)	
Prevention					<.001
Baseline	16.6 (24.0)	14.0 (22.4)	2.5 (−3.3 to 8.3)	2.8 (−2.9 to 8.5)	
3 months	55.4 (22.7)	14.6 (21.6)	40.7 (34.8 to 46.5)	40.8 (35.1 to 46.6)	
12 months	50.1 (26.4)	13.7 (25.5)	36.3 (30.2 to 42.5)	36.5 (30.5 to 42.6)	

Table 4. Comparison of knowledge about the 3 target diseases, shown as mean percentage of correct answers, between participants who used the information spots (InfoSpots) and those who did not at the 12-month assessment.

Target disease and time period	InfoSpots users (n=73), mean (SD)	Non–InfoSpots users (n=170), mean (SD)
HIV/AIDS		
3 months	55.1 (18.0)	53.1 (16.0)
12 months	63.6 (11.1)	56.8 (11.6)
Tuberculosis		
3 months	34.8 (16.5)	33.9 (15.1)
12 months	41.1 (11.4)	33.6 (10.8)
Taenia solium **cysticercosis and taeniosis**		
3 months	38.5 (20.3)	39.9 (20.3)
12 months	50.6 (17.6)	36.7 (21.8)

Discussion

Principal Findings

The DigI study is the first of its kind to assess health knowledge uptake and retention following a 2-pronged digital health education intervention consisting of watching health animations and accessing a free digital health education platform (InfoSpots) in rural Tanzania.

Participants in the intervention group attained significantly higher levels of knowledge than those in the control group, with the adjusted differences among the groups being 10.2%, 12%, and 31.5% for HIV/AIDS, TB, and TSCT, respectively, after 12 months. We observed increased knowledge of all 3 diseases in all domains. The effect on knowledge retention was particularly strong for TSCT, with a low level of knowledge at baseline, as shown by the increase from 9.7% to 42.1% after 12 months.

Looking at the domains, the effect was strongest in the 3-month assessment for all TB and 3 out of the 4 TSCT domains. In contrast, for all HIV/AIDS and TSCT symptoms domains, an increased effect was observed at the 12-month assessment, indicating that the use of InfoSpots for health education may have contributed to the retention and enhancement of health knowledge. In fact, the InfoSpots users showed further increase in knowledge at 12 months compared with those who did not use InfoSpots. This increase in knowledge was more pronounced for HIV/AIDS and TSCT than for TB. However, this does not exclude the possibility that the participants accessed other knowledge sources over the course of the study. This may also explain the slight increase in knowledge in the control group at 12 months, which was 1.9%, 6.8%, and 4.4% for HIV/AIDS, TB, and TSCT, respectively. We were unable to assess why the participants in the control group improved their knowledge. They were not asked any questions about whether they had talked to other villagers about diseases or had otherwise obtained information. Contamination among the groups is unlikely, given the long distance of at least two and a half hours by car. Between the baseline and the 12-month assessment, 107 out of 600 participants dropped out, accounting for 17.8% of the initial participants. All participants were contacted by subvillage leaders via their mobile phone number. The reasons for dropping out were, in most cases, migration for work opportunities in other villages, as the data collection lasted for more than a year

in an area characterized by seasonal workers in agriculture, fishing activities, and petty trading.

The intervention effect was independent of age, sex, education, and occupation and demonstrated that digital health education is applicable and useful for all people despite sociodemographic differences.

Comparison With Prior Work

The DigI study shows that animations with short, clear, and basic health messages have significant potential to improve health knowledge in the short to long term, as confirmed by previous studies in a sub-Saharan setting [14-18]. Despite challenges in comparing our results directly with other intervention studies using animations for health knowledge transfer—owing to differences in health promotion content, interventions, target groups, and research design, 3 studies are worth mentioning. Adam et al [18] observed a small significant increase in knowledge among pregnant women in a month, but not 5 months, after their video breastfeeding intervention. Lund et al [16] found a 38% increase in maternal health knowledge among health care workers in their intervention study 12 months after introducing a safe delivery app with animations for health knowledge improvement, whereas Bolan et al [17] researched the same app and found a 17% to 19.5% increase in their 3-month assessment. Although health care workers may have some prior medical knowledge, the results indicate that the study population and setting, and most probably other factors, may impact knowledge uptake and retention and always need to be considered.

Strengths and Limitations

Overall, the DigI study, which offers digital health education both directly and via local InfoSpots, is unique and novel and could inform the rethinking of health education delivery in rural areas of low- and middle-income countries as a whole. It seems that simple animations co-designed with local stakeholders, provided in Swahili with local characters in a recognizable environment [25], can contribute to improvement in health knowledge. The repetition of key messages in animations and illustrative scenes accompanied by the narrator's explanations (Figure 2) could be key factors for better knowledge uptake and retention.

Community health workers, equipped with a digital tool, could share tailored health messages while simultaneously sensitizing individuals to the use of a locally adapted digital health promotion platform that can be deployed at strategic locations within the communities. In fact, we were able to show that InfoSpots users were creative and enthusiastic when using the platform and that the animations were well accepted as a knowledge transfer tool in the communities [31]. Although InfoSpots require cross-sectoral collaboration among community leaders, internet providers, and the health and educational sectors, if co-designed and promoted together, they have the potential to create ownership and thus achieve long-term sustainability. Furthermore, such InfoSpots can contribute to the development of digital skills and provide people with access to essential information, thus supporting the achievement of SDG 3, SDG 4, and SDG 16.

The DigI study has some limitations that require further investigation. First, the selection of intervention and control villages was nonrandomized and performed for technical reasons related to the installation of the InfoSpots. The demographic data showed differences between the intervention and control groups in terms of age, sex, education, and occupation. This may have affected the external validity of the DigI study's research design. However, our adjusted results showed that none of the possible confounding factors affected the intervention effect. Participants were randomly selected from all villages and subvillages, which is a strength that reduces the selection bias. Furthermore, with this study design, it was not possible to assess whether improved health knowledge and resulting health literacy may have an impact on the burden of disease from our 3 target diseases in the communities. Increased health knowledge does not guarantee a change in health behaviors. To initiate and sustain behavior change, multiple interventions at various levels are required [1], and this project only focused on interventions that improve health knowledge.

Future Directions

In summary, the DigI study showed that the unique combination of animated health videos shown to participants in their homes and the use of free InfoSpots with an integrated digital health platform improved knowledge about the 3 target diseases, not only in the short term but also over a longer time frame. Furthermore, implementing health access points such as InfoSpots in rural Tanzania, with animations and other digital health education materials in their own language, may contribute to achieving SDG 3 and in our case, particularly target SDG 3.3—ending the epidemics of HIV/AIDS, TB, and neglected tropical diseases by 2030. However, further studies are required to demonstrate the impact of the promising digital health intervention of the DigI study on health-seeking behavior and thus on disease prevention and management.

Acknowledgments

The DigI project and its DigI study were funded by the Norwegian Research Council (project number 267558) and Norwegian Ministry of Foreign Affairs and Norwegian Agency for Development Cooperation. The authors would like to thank the members of the DigI project and the participants in Migoli, Izazi, Kimande, and Idodi for their contribution to the DigI study.

Data Availability

Deidentified participant-level data and the data dictionary will be made available on request, with a signed data access agreement after December 2022. Please email the corresponding author for their access.

Authors' Contributions

CH, FS, BN, LMD, JN, and ASW contributed to study concept and design. CH collected the data. CH, DS, and LMD analyzed and interpreted the data. CH and DS drafted the manuscript. All the authors critically revised the manuscript for important intellectual content. DS and LMD performed statistical analyses. JN obtained funding. CH, FS, BN, JN, and ASW provided the administrative, technical, and material support. BN, FS, and ASW supervised this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
HIV/AIDS health animated video (English).
[MP4 File (MP4 Video), 58029 KB - jmir_v24i7e37666_app1.mp4]

Multimedia Appendix 2
Tuberculosis health animated video (English).
[MP4 File (MP4 Video), 28571 KB - jmir_v24i7e37666_app2.mp4]

Multimedia Appendix 3
Taenia solium cysticercosis and taeniosis health animated video (English).
[MOV File , 78062 KB - jmir_v24i7e37666_app3.mov]

Multimedia Appendix 4
Study questionnaire (HIV, tuberculosis, or Taenia solium cysticercosis and taeniosis knowledge questions used at baseline and at 3 and 12 months).
[PDF File (Adobe PDF File), 398 KB - jmir_v24i7e37666_app4.pdf]

References

1. Green L, Kreuter M, Deeds S, Partridge K. Health Education Planning: A Diagnostic Approach. U.S: Brown (William C.); 1980.
2. Education for Health: A Manual on Health Education in Primary Health Care. Geneva: World Health Organization; 1988.
3. Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, (HLS-EU) Consortium Health Literacy Project European. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012 Jan 25;12:80 [FREE Full text] [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]
4. The mandate for health literacy. World Health Organization. URL: https://www.who.int/teams/health-promotion/enhanced-wellbeing/ninth-global-conference/health-literacy [accessed 2022-06-08]
5. Gellert P, Tille F. What do we know so far? The role of health knowledge within theories of health literacy. Eur Health Psychologist 2015;17:266-274.
6. Trevethan R. Deconstructing and assessing knowledge and awareness in public health research. Front Public Health 2017;5:194 [FREE Full text] [doi: 10.3389/fpubh.2017.00194] [Medline: 28824900]
7. Regional factsheets. Africa – eastern and southern 2020. UNAIDS. URL: https://www.unaids.org/sites/default/files/media_asset/UNAIDS_FactSheet_en.pdf [accessed 2022-06-08]
8. Tuberculosis. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/tuberculosis [accessed 2022-06-08]
9. Taeniasis/cysticercosis. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/taeniasis-cysticercosis [accessed 2022-06-08]
10. Communications statistics - Quarter 2 - 2021/2022. TCRA Tanzania. URL: https://www.tcra.go.tz/uploads/text-editor/files/Communications%20Statistics%20December%202021%20(1)_1645554800.pdf [accessed 2022-06-08]
11. The World Bank in Tanzania. The World Bank. URL: https://www.worldbank.org/en/country/tanzania/overview#1 [accessed 2022-06-08]
12. McHenry MS, Fischer LJ, Chun Y, Vreeman RC. A systematic review of portable electronic technology for health education in resource-limited settings. Glob Health Promot 2019 Jun;26(2):70-81. [doi: 10.1177/1757975917715035] [Medline: 28832243]
13. Adam M, McMahon SA, Prober C, Bärnighausen T. Human-centered design of video-based health education: an iterative, collaborative, community-based approach. J Med Internet Res 2019 Jan 30;21(1):e12128 [FREE Full text] [doi: 10.2196/12128] [Medline: 30698531]
14. Bello-Bravo J, Dannon E, Agunbiade T, Tamo M, Pittendrigh B. The prospect of animated videos in agriculture and health: a case study in Benin. Int J Educ Develop Inf Commun Technol 2013;9(3):4-16 [FREE Full text]

15. Hébert C, Dagenais C, Mc Sween-Cadieux E, Ridde V. Video as a public health knowledge transfer tool in Burkina Faso: a mixed evaluation comparing three narrative genres. PLoS Negl Trop Dis 2020 Jun;14(6):e0008305 [FREE Full text] [doi: 10.1371/journal.pntd.0008305] [Medline: 32520930]
16. Lund S, Boas IM, Bedesa T, Fekede W, Nielsen HS, Sørensen BL. Association between the safe delivery app and quality of care and perinatal survival in Ethiopia: a randomized clinical trial. JAMA Pediatr 2016 Aug 01;170(8):765-771. [doi: 10.1001/jamapediatrics.2016.0687] [Medline: 27322089]
17. Bolan NE, Sthreshley L, Ngoy B, Ledy F, Ntayingi M, Makasy D, et al. mLearning in the democratic republic of the Congo: a mixed-methods feasibility and pilot cluster randomized trial using the safe delivery app. Glob Health Sci Pract 2018 Dec 27;6(4):693-710 [FREE Full text] [doi: 10.9745/GHSP-D-18-00275] [Medline: 30591577]
18. Adam M, Johnston J, Job N, Dronavalli M, Le Roux I, Mbewu N, et al. Evaluation of a community-based mobile video breastfeeding intervention in Khayelitsha, South Africa: the Philani MOVIE cluster-randomized controlled trial. PLoS Med 2021 Sep;18(9):e1003744 [FREE Full text] [doi: 10.1371/journal.pmed.1003744] [Medline: 34582438]
19. Holst C, Sukums F, Ngowi B, Diep LM, Kebede TA, Noll J, et al. Digital health intervention to increase health knowledge related to diseases of high public health concern in Iringa, Tanzania: protocol for a mixed methods study. JMIR Res Protoc 2021 Apr 22;10(4):e25128 [FREE Full text] [doi: 10.2196/25128] [Medline: 33885369]
20. Tanzania HIV impact survey (THIS) 2016-2017. Tanzania Commission for AIDS, Zanzibar AIDS Commission. URL: https://www.nbs.go.tz/nbs/takwimu/this2016-17/Tanzania_SummarySheet_English.pdf [accessed 2021-12-13]
21. TB prevalence in Tanzania. The National Tuberculosis and Leprosy Programme (NTLP). URL: https://ntlp.go.tz/tuberculosis/tb-prevalence-in-tanzania/ [accessed 2022-06-08]
22. Ngowi HA, Winkler AS, Braae UC, Mdegela RH, Mkupasi EM, Kabululu ML, et al. Taenia solium taeniosis and cysticercosis literature in Tanzania provides research evidence justification for control: a systematic scoping review. PLoS One 2019;14(6):e0217420 [FREE Full text] [doi: 10.1371/journal.pone.0217420] [Medline: 31166983]
23. Updated environmental and social impact assessment (ESIA) study for the proposed 292km Iringa-Mbeya 400kv power transmission and associated Subsations at Kisada (Mafinga) and Iganjo (Mbeya) in Iringa, Njombe and Mbeya regions, Tanzania. Tanzania Electric Supply Company. URL: http://documents1.worldbank.org/curated/pt/808631519143051939/pdf/SFG4014-V2-EA-P163752-PUBLIC-Disclosed-2-20-2018.pdf [accessed 2022-06-08]
24. Lavrakas P. Kish selection method. In: Encyclopedia of Survey Research Methods. Thousand Oaks, California: SAGE Publications; 2008.
25. Holst C, Isabwe GM, Sukums F, Ngowi H, Kajuna F, Radovanović D, et al. Development of digital health messages for rural populations in Tanzania: multi- and interdisciplinary approach. JMIR Mhealth Uhealth 2021 Sep 22;9(9):e25558 [FREE Full text] [doi: 10.2196/25558] [Medline: 34550081]
26. Digital health video download. University of Oslo. URL: https://www.med.uio.no/helsam/english/research/centres/global-health/knowledge-resources/digital-health/digital-health-videos/ [accessed 2022-06-08]
27. Non-discriminating access for Digital Inclusion (DigI project). Migoli Village homepage - Health information. URL: https://migoli.yeboo.com/health_information_dashboard.php [accessed 2022-06-08]
28. Mwidunda SA, Carabin H, Matuja WB, Winkler AS, Ngowi HA. A school based cluster randomised health education intervention trial for improving knowledge and attitudes related to Taenia Solium cysticercosis and Taeniasis in Mbulu district, northern Tanzania. PLoS One 2015;10(2):e0118541 [FREE Full text] [doi: 10.1371/journal.pone.0118541] [Medline: 25719902]
29. Mutayoba B, Ngowi B, Kohi W. HIV prevalence and related risk factors in prison settings: findings from a rapid situational assessment in mainland Tanzania. In: Proceedings of the 20th International AIDS Conference. 2014 Presented at: 20th International AIDS Conference; Jul 12-21, 2014; Melbourne.
30. Stata Statistical Software: Release 16. College Station, TX: StataCorp LP; 2019.
31. Holst C, Tschirhart N, Ngowi B, Noll J, Winkler AS. Utilizing community InfoSpots for health education: perspectives and experiences in Migoli and Izazi, Tanzania. Health Promot Int 2021 Dec 13 (forthcoming). [doi: 10.1093/heapro/daab187] [Medline: 34897433]

Abbreviations

InfoSpots: information spots
SDG: Sustainable Development Goal
TB: tuberculosis
TSCT: Taenia solium cysticercosis and taeniosis

Edited by T Leung; submitted 09.03.22; peer-reviewed by A Garcia, R Halkes; comments to author 02.06.22; revised version received 09.06.22; accepted 16.06.22; published 28.07.22.

Please cite as:
Holst C, Stelzle D, Diep LM, Sukums F, Ngowi B, Noll J, Winkler AS
Improving Health Knowledge Through Provision of Free Digital Health Education to Rural Communities in Iringa, Tanzania: Nonrandomized Intervention Study
J Med Internet Res 2022;24(7):e37666
URL: https://www.jmir.org/2022/7/e37666
doi:10.2196/37666
PMID:35900820

Original Paper

Mobile Health Technology Use and the Acceptability of an mHealth Platform for HIV Prevention Among Men Who Have Sex With Men in Malaysia: Cross-sectional Respondent-Driven Sampling Survey

Roman Shrestha[1], MPH, PhD; Francesca Maviglia[2], MPH; Frederick L Altice[2], MD; Elizabeth DiDomizio[2], MPH; Antoine Khati[1], MD; Colleen Mistler[1], MSc; Iskandar Azwa[3], MD; Adeeba Kamarulzaman[3], MD; Mohd Akbar Ab Halim[3], BA; Jeffrey A Wickersham[2], PhD

[1]Department of Allied Health Sciences, University of Connecticut, Storrs, CT, United States

[2]Section of Infectious Diseases, Yale University, New Haven, CT, United States

[3]Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia

Corresponding Author:
Roman Shrestha, MPH, PhD
Department of Allied Health Sciences
University of Connecticut
358 Mansfield Rd
Storrs, CT, 06269
United States
Phone: 1 860 486 2446
Email: roman.shrestha@uconn.edu

Abstract

Background: The growth in mobile technology access, utilization, and services holds great promise in facilitating HIV prevention efforts through mobile health (mHealth) interventions in Malaysia. Despite these promising trends, there is a dearth of evidence on the use of mHealth platforms that addresses HIV prevention among Malaysian men who have sex with men.

Objective: The goal of this study was to gain insight into (1) access and utilization of communication technology (eg, landline phone, internet, mobile phone), (2) acceptability of mHealth-based interventions for HIV prevention services, and (3) preferences regarding the format and frequency of mHealth interventions among Malaysian men who have sex with men.

Methods: We conducted a cross-sectional survey with Malaysian men who have sex with men between July 2018 and March 2020. Participants were recruited using respondent-driven sampling in the Greater Kuala Lumpur region of Malaysia. We collected information on demographic characteristics, HIV risk-related behaviors, access to and the frequency of use of communication technology, and acceptability of using mHealth for HIV prevention using a self-administered questionnaire with a 5-point scale (1, never; 2, rarely; 3, sometimes; 4, often; 5, all the time).

Results: A total of 376 men participated in the survey. Almost all respondents owned or had access to a smartphone with internet access (368/376, 97.9%) and accessed the internet daily (373/376, 99.2%), mainly on a smartphone (334/376, 88.8%). Participants on average used smartphones primarily for social networking (mean 4.5, SD 0.8), followed by sending or receiving emails (mean 4.0, SD 1.0), and searching for health-related information (mean 3.5, SD 0.9). There was high acceptance of the use of mHealth for HIV prevention (mean 4.1, SD 1.5), including for receiving HIV prevention information (345/376, 91.8%), receiving medication reminders (336/376, 89.4%), screening and monitoring sexual activity (306/376, 81.4%) or illicit drug use (281/376, 74.7%), and monitoring drug cravings (280/376, 74.5%). Participants overwhelmingly preferred a smartphone app over other modalities (eg, text, phone call, email) for engaging in mHealth HIV prevention tools. Preference for app notifications ranged from 186/336 (53.9%), for receiving HIV prevention information, to 212/336 (69.3%), for screening and monitoring sexual activity. Acceptance of mHealth was higher for those who were university graduates (P=.003), living in a relationship with a partner (P=.04), engaged in sexualized drug use (P=.01), and engaged in receptive anal sex (P=.006).

Conclusions: Findings from this study provide support for developing and deploying mHealth strategies for HIV prevention using a smartphone app in men who have sex with men—a key population with suboptimal engagement in HIV prevention and treatment.

(J Med Internet Res 2022;24(7):e36917) doi:10.2196/36917

KEYWORDS

HIV; mHealth; men who have sex with men; mobile phone; Malaysia; mobile health; HIV prevention; sexual health; public health; digital health; communication technology; health technology; technology accessibility; smartphone app; HIV treatment

Introduction

In comparison with high-income countries, low- and middle-income countries have disproportionately high rates of HIV because funding and access to HIV prevention and treatment services are limited [1]. Men who have sex with men are at increased vulnerability to HIV due to a combination of biological factors (ie, sexual exposure risk, concurrent sexually transmitted infections) [2] and a higher sexual network or community HIV prevalence [3,4]. In addition, behavioral factors (eg, condomless sex, multiple sexual partners, sexualized drug use which is also known as *Chemsex*) have also been found to increase the risk of HIV in this group [4,5].

In Malaysia, HIV prevalence among men who have sex with men has grown exponentially in the past decade, peaking at 21.6% in 2017, compared with 0.4% among the general population [6,7]. The causes of Malaysia's expanding HIV epidemic among men who have sex with men are multifactorial. In populations such as men who have sex with men, the uptake of HIV prevention and treatment services is often low due to reluctance to disclose their sexual orientation, lack of anonymity, and concerns about confidentiality [8]. In Malaysia, which is a middle-income country with a Muslim majority, same-sex sexual behavior is a crime according to both secular and Sharia criminal laws, which contributes to high levels of stigma and discrimination. The criminalization of same-sex relationships has been found to be associated with lower access to condoms, lubricants, HIV testing, and HIV treatment in a study [9] with a sample of men who have sex with men from 115 countries. Several studies [10-13] have documented negative attitudes and discrimination toward men who have sex with men in Malaysia, including in health care settings. Consequently, the scale-up of evidence-based HIV testing, prevention, and treatment programs that target those who are most vulnerable to HIV, including men who have sex with men, has been negatively impacted. Low HIV-testing uptake among men who have sex with men in Malaysia has resulted in many men who have sex with men being diagnosed with HIV at an advanced stage [8,14-16]. There is, therefore, a need for innovative strategies to reach with services and information and retain into care stigmatized populations such as men who have sex with men in Malaysia.

The use of mobile health (mHealth) technology to improve health outcomes has been expanding globally, and research is needed to guide its use in health care delivery [17-23]. mHealth has been found to be a promising and cost-effective strategy to reach and serve stigmatized and hidden populations such as men who have sex with men [24-27]. In Malaysia, leveraging an mHealth platform may allow users to feel safer and less vulnerable to potential legal or social harm (for example, by reducing face-to-face interactions with providers) and bypass barriers to care for marginalized populations [28-32]. Over the past decade, the use of mobile technology in Malaysia has grown markedly, with a mobile phone penetration rate of 97.5% and an internet penetration rate of 71.1% [33]. With increases in access to, utilization of, and services using communication technologies, an mHealth approach holds great promise to facilitate HIV prevention efforts in Malaysia.

Despite these promising trends, there is a dearth of evidence on the use of mHealth platforms to address HIV prevention in Malaysian men who have sex with men. Thus, the goal of this study was to gain insight into (1) access to and utilization of communication technology (eg, landline phone, internet, cell phone), (2) acceptability of mHealth-based interventions for HIV prevention services, and (3) preferences regarding the format and frequency of mHealth interventions among Malaysian men who have sex with men.

Methods

Study Design and Setting

We conducted a cross-sectional study with participants recruited from the Greater Kuala Lumpur region between July 2018 and March 2020. Data collection was conducted in a private room at the Centre of Excellence for Research in AIDS (University of Malaya), because of its central location in the Greater Kuala Lumpur region, accessibility by public transport, and ability to provide free parking to participants.

Study Participants and Procedures

Inclusion criteria were (1) being 18 years or older; (2) identifying as a cisgender male; (3) currently residing in the Greater Kuala Lumpur region; (4) reporting having been engaged in sexual activity with male sexual partners in the previous 6 months; and (5) being able to read and understand English, Chinese, or Bahasa Malaysia.

The questionnaires were created in English and then translated to Chinese and Bahasa Malaysia. We used forward-backward translation and pilot tested the translated questionnaires to ensure translation quality.

Participants were recruited using respondent-driven sampling, which is a network-based sampling method that is often used for hard-to-reach populations and combines peer-driven recruitment and statistical adjustments to reduce bias and approximate random sampling [34]. During our formative work, initial respondent-driven sampling participants, called seed participants, were carefully selected with assistance from community-based organizations for men who have sex with men. We aimed to recruit a seed participant sample (n=27) that reflected the diversity (eg, ethnicity, age) of the community of men who have sex with men in the region.

Each participant who completed the study was given 3 coupons to recruit men from their peer networks, who were, in turn, also given 3 coupons. Each coupon card contained a unique

respondent-driven sampling number that allowed us to trace the peer recruitment chain and important study-related information (study site address, contact information for the study team, and inclusion criteria). Coupon management software was used to track distributed and redeemed coupons during the study, and a standard numbering system was used to track the recruiter-recruit relationship. Participants received 30 Malaysian Ringgit (at the time of publication, 1 MYR was approximately equivalent to US $0.23) for study participation and an additional 10 MYR for each peer who was successfully recruited to the study (up to 3 peers).

Individuals could choose whether to come during walk-in hours or set an appointment by phone. The research site was open 7 days a week to offer maximum flexibility and accommodate different work schedules. Individuals who presented with a valid coupon underwent initial eligibility screening. If eligible, they were asked to provide informed consent. Each participant completed the web-based questionnaire (Qualtrics, Qualtrics XM) in a private room, which took approximately 20 minutes, while study staff waited outside the room, to ensure privacy.

Ethics Approval

The study was approved by the institutional review boards of the University of Malaya (201854) and Yale University (2000023152).

Measures

Participant Characteristics

We collected participant characteristics (age, ethnicity, educational status, relationship status, income, history of childhood physical abuse, history of childhood sexual assault, and depressive symptoms experienced in the past week).

Access to and Frequency of Use of Communication Technology

We adapted a scale that we used in previous studies [35-37] to measure participants' access to and frequency of use of various types of communication devices, including landline telephone, mobile phone with internet access (smartphone) and without internet access (basic phone), tablet, laptop, and personal computer. Participants were asked how often they use each technology on a 5-point Likert scale (ranging from 1, never, to 5, all the time).

Additionally, participants were asked if they had daily access to the internet, which device was their primary device for accessing the internet, and the number of hours spent on the internet each week. Participants' utilization of smartphones for various internet-based activities (including social networking, sending or receiving emails, using geosocial networking apps or websites, searching for health-related information, or using health-related apps) was assessed using a 5-point Likert scale (ranging from 1, never, to 5, all the time). Participants were also asked which men who have sex with men–related geosocial networking apps or websites (eg, Grindr, Hornet, Planet Romeo) they currently used.

Acceptability of mHealth

The mHealth acceptance scale was adapted from previous studies [35-37]. Participants were asked about their willingness to use 5 mHealth-related features—receiving medication reminders, monitoring drug cravings, screening and monitoring illicit drug use, screening and monitoring sexual activity, and receiving information about HIV prevention. Each feature was rated on a 5-point Likert scale (ranging from 1, not willing, to 5, extremely willing); the scale was later dichotomized for analysis, with "not willing" coded as *no* and "somewhat willing," "willing," "very willing," "extremely willing" coded as *yes*. An mHealth acceptance score was created by taking an average cumulative score of the 5 mHealth-related dichotomized variables, with a higher score indicating greater acceptance (α=.88). Respondents' preferred frequency (ie, daily, weekly, and monthly) and modality of mHealth (ie, phone call, text message, email, or app) were also assessed [35-37].

Childhood Trauma and Mental Health

Two items from the US Centers for Diseases Control and Prevention's Behavioral Risk Factor Surveillance System questionnaire [38] were used to measure history of childhood and physical and sexual trauma. Childhood physical trauma was measured with a single-item question, "Before the age of 18, were you ever hit, slapped, kicked, or physically hurt by an adult?" Childhood sexual trauma was measured with 2 items: "Before the age of 18, were you ever forced to have sex by an adult or older child?" and "Before the age of 18, were you ever touched in a sexual way by an adult or older child when you did not want to be touched that way or were you ever forced to touch an adult or older child in a sexual way?" A "yes" response to either question resulted in a *yes* coding for the presence of childhood sexual trauma. Depressive symptoms were assessed using the 10-item Center for Epidemiological Studies Depression scale [39,40]. The total sum score ranges from 0 to 30, with a standard cut-off (score >10) for moderate to severe depression (α=.89) [40].

Sexual and Drug-Related Behaviors

Participants were asked information about their sexual behavior, including recent (past 6 months) engagement in anal sex and in which role (ie, insertive or receptive; participants were able to select both roles if applicable); recent engagement in an HIV-serodiscordant sexual relationship; consistent condom use; and lifetime engagement in sexualized drug use, which we defined as any use of crystal methamphetamine, gamma-hydroxybutyrate, gamma-butyrolactone, or 5-methoxy-N, N-diisopropyl tryptamine (commonly known as foxy or foxy methoxy [41]) before or during sexual activity. Additionally, participants were asked about any lifetime injection drug use. The 6-month cut-off point for sexual activity and engagement in a serodiscordant relationship was chosen based on the Centers for Disease Control and Prevention guidelines [42].

Participants were asked if they had ever been tested for or diagnosed with HIV or other sexually transmitted infections, including chlamydia (*Chlamydia trachomatis*), gonorrhea (*Neisseria gonorrhoeae*), syphilis (*Treponema pallidum*),

hepatitis B (*Orthohepadnavirus hepatitis B virus*), and hepatitis C (*Hepacivirus hepacivirus C*), and if they had ever used pre-exposure prophylaxis or postexposure prophylaxis for HIV prevention.

Data Analysis

Analyses were performed using SPSS software (version 26; IBM Corp). We calculated descriptive statistics, such as frequencies and percentages for categorical variables and means and standard deviations for continuous variables, and used multivariate linear regression to assess factors associated with the primary outcome—willingness to use mHealth (measured by the mHealth acceptability scale (continuous variable). Candidate covariates were selected based on previous literature on mHealth acceptability and were included in the multivariable model if $P<.05$ in a bivariate model. Estimates were evaluated for statistical significance based on probability criteria of $P<.05$.

Results

Participant Characteristics

A total of 376 men (age: mean 27.5 years, SD 6.5 years) participated in the survey. Most participants identified ethnically as Malay (220/376, 58.5%). Over half of the participants were university graduates (216/376, 57.4%) and single (216/376, 56.4%). The mean monthly income was 3602.9 MYR. The majority of participants (222/376, 59.0%) reported symptoms consistent with moderate to severe depression (Table 1).

Although almost all participants had engaged in anal sex in the past 6 months (363/376, 96.5%; receptive role: 285/376, 75.8%), only one-fifth (72/376, 19.1%) reported consistent condom use. Moreover, one-fifth of participants (82/376, 21.8%) reported having ever engaged in sexualized drug use. Overall, 71.0% (267/376) of the participants had been tested for HIV at least once in their lifetime, and 27.4% (103/376) had been previously diagnosed with a sexually transmitted infection other than HIV. Only a small proportion of participants had ever used pre-exposure prophylaxis (26/376, 6.9%) or postexposure prophylaxis (27/376, 7.2%).

Table 1. Characteristics of participants.

Characteristic	Respondents (n=376)
Age (years), mean (SD)	27.5 (6.5)
Ethnicity (Malaya), n (%)	
No	156 (41.5)
Yes	220 (58.5)
University graduate[a], n (%)	
No	160 (42.6)
Yes	216 (57.4)
Relationship status, n (%)	
Single	212 (56.4)
Partner	164 (43.6)
Monthly income (MYR)[b], mean (SD)	3602.9 (5082.6)
Ever had HIV test, n (%)	
No	109 (29.0)
Yes	267 (71.0)
Previously diagnosed with STI[c], n (%)	
No	273 (72.6)
Yes	103 (27.4)
Ever used pre-exposure prophylaxis, n (%)	
No	350 (93.1)
Yes	26 (6.9)
Ever used postexposure prophylaxis, n (%)	
No	349 (92.8)
Yes	27 (7.2)
Experienced childhood physical abuse, n (%)	
No	199 (52.9)
Yes	177 (47.1)
Experienced childhood sexual assault, n (%)	
No	235 (62.5)
Yes	141 (37.5)
Depressive symptoms, n (%)	
No	154 (41.0)
Yes	222 (59.0)
Ever injected drugs, n (%)	
No	359 (95.5)
Yes	17 (4.5)
Engaged in anal sex (past 6 months), n (%)	
No	13 (3.5)
Yes	363 (96.5)
Type of anal sex (past 6 months)[d], n (%)	
Insertive	271 (72.1)
Receptive	285 (75.8)

Characteristic	Respondents (n=376)
HIV-serodiscordant relationship (past 6 months), n (%)	
No	340 (90.4)
Yes	36 (9.6)
Consistent condom use (past 6 months), n (%)	
No	304 (80.9)
Yes	72 (19.1)
Ever engaged in sexualized drug use, n (%)	
No	294 (78.2)
Yes	82 (21.8)

[a]This category included college, university, or professional degrees.

[b]MYR: Malaysian Ringgit (1 MYR is approximately US $0.23).

[c]STI: sexually transmitted infections.

[d]The total exceeds 100% because the options were not mutually exclusive.

Access to and Frequency of Use of Communication Technology

Almost all participants (368/376, 97.9%) owned or had access to a smartphone with internet access, and 13.3% (50/376) of participants had access to a basic cell phone without internet access (Table 2). More than two-thirds of participants (270/376, 71.8%) reported having access to a laptop; between one-quarter and one-fifth of participants had access to a personal computer (100/376, 26.6%), a tablet (85/376, 22.6%), and a landline telephone (81/376, 21.6%). The frequency of use of each device was largely consistent with the frequency of ownership and access, with smartphones representing the most frequently used technology (mean 4.9, SD 0.4), followed by laptops (mean 3.8, SD 1.2), personal computers (mean 2.5, SD 1.4), tablets (mean 2.3, SD 1.3), basic cell phones (mean 2.1, SD 1.3), and landline telephones (mean 1.8, SD 0.9).

Almost all participants (373/376, 99.2%) accessed the internet daily, largely through a smartphone (334/376, 88.8%), and spent on average 9.4 hours per week (SD 4.9) on the internet (Table 3). The most common activities that participants used the internet on their smartphones for were social networking (mean 4.5, SD 0.8) and sending or receiving emails (mean 4.0, SD 1.0). Participants also used their smartphones to access geosocial networking apps or websites (mean 3.6, SD 1.1), search for health-related information (mean 3.5, SD 0.9), and use health-related apps (mean 2.9, SD 1.1). The majority (345/376, 91.8%) of participants used geosocial networking apps, with Grindr, Blued, and Hornet being the most popular.

Table 2. Ownership or access to and frequency of use of communication technology.

Variable	Ownership or access (n=376), n (%)	Frequency of use[a], mean (SD)
Mobile phone		
With internet access (smartphone)	368 (97.9)	4.9 (0.4)
Without internet access (basic phone)	50 (13.3)	2.1 (1.3)
Laptop	270 (71.8)	3.8 (1.2)
Personal computer	100 (26.6)	2.5 (1.4)
Tablet	85 (22.6)	2.3 (1.3)
Landline telephone	81 (21.5)	1.8 (0.9)

[a]This was assessed using a 5-point Likert scale (1, never; 2, rarely; 3, sometimes; 4, often; 5, all the time).

Table 3. Access to internet.

Variables	Respondents (n=376)
Daily access to the internet, n (%)	
No	3 (0.8)
Yes	373 (99.2)
Primary device for accessing the internet, n (%)	
Smartphone	334 (88.8)
Laptop	21 (5.6)
Personal computer	6 (1.6)
Others	15 (4.0)
Time spent on the internet (hours per week), mean (SD)	9.4 (4.9)
Use of the internet on a smartphone for various activities[a], mean (SD)	
Online social networking	4.5 (0.8)
Send or receive emails	4.0 (1.0)
Geosocial networking apps or websites	3.6 (1.1)
Search for health-related information	3.5 (0.9)
Use health-related apps	2.9 (1.1)

[a]This item was assessed using a 5-point Likert scale (1, never; 2, rarely; 3, sometimes; 4, often; 5, all the time).

Acceptability of mHealth

The majority of participants were interested in receiving HIV prevention information (345/376, 91.8%) on a monthly (147/376, 42.6% of those who expressed willingness) or weekly (131/376, 38.0%) basis, and in receiving medication reminders (336/376, 89.4%) mostly on a daily basis (191/376, 56.8%). Additionally, there was interest in using mHealth to screen and monitor sexual activity (306/376, 81.4%) on a weekly (135/376, 44.1%) or monthly (99/376, 32.4%) basis; screen and monitor illicit drug use (281/376, 74.7%) on a weekly (104/376, 37.0%) or monthly (100/376, 35.6%) basis; and monitor drug cravings (280/376, 74.5%) on a weekly (115/376, 41.1%) or monthly (79/376, 28.2%) basis. The preferred modality of mHealth strategies was via apps, regardless of the type of intervention (Table 4).

Table 4. Interest in and acceptance of mobile health (mHealth) among participants (N=376).

Interest in using mHealth to...	No, n (%)	Yes, n (%)
Receive medication reminders	40 (10.6)	336 (89.4)
Preferred frequency (n=374)		
Daily	13 (32.5)	191 (56.8)
Weekly	10 (25.0)	91 (27.1)
Monthly	8 (20.0)	44 (13.1)
Never	8 (20.0)	9 (2.7)
Preferred mechanism (n=375)		
Phone calls	10 (25.0)	15 (4.5)
Text messages	13 (32.5)	93 (27.7)
App notification	12 (30.0)	206 (61.3)
Email	4 (10.0)	22 (6.5)
Monitor drug cravings	96 (25.5)	280 (74.5)
Preferred frequency (n=375)		
Daily	10 (10.4)	48 (17.1)
Weekly	17 (17.7)	115 (41.1)
Monthly	15 (15.6)	79 (28.2)
Never	53 (55.2)	38 (13.6)
Preferred mechanism (n=375)		
Phone calls	11 (11.5)	2 (0.7)
Text messages	23 (24.0)	65 (23.2)
App	45 (46.9)	183 (65.4)
Email	16 (16.7)	30 (10.7)
Screen and monitor illicit drug use	95 (25.3)	281 (74.7)
Preferred frequency (n=375)		
Daily	10 (10.5)	50 (17.8)
Weekly	17 (17.9)	104 (37.0)
Monthly	17 (17.9)	100 (35.6)
Never	50 (52.6)	27 (9.6)
Preferred mechanism (n=375)		
Phone calls	9 (9.5)	5 (1.8)
Text messages	22 (23.2)	55 (19.6)
App	45 (47.4)	189 (67.3)
Email	18 (18.9)	32 (11.4)
Screen and monitor sexual activity	70 (18.6)	306 (81.4)
Preferred frequency (n=373)		
Daily	10 (14.3)	58 (19)
Weekly	15 (21.4)	135 (44.1)
Monthly	17 (24.3)	99 (32.4)
Never	26 (37.1)	13 (4.2)
Preferred mechanism (n=375)		
Phone calls	8 (11.4)	5 (1.6)
Text messages	15 (21.4)	59 (19.3)

Interest in using mHealth to...	No, n (%)	Yes, n (%)
App	32 (45.7)	212 (69.3)
Email	14 (20.0)	30 (9.8)
Receive HIV prevention information	31 (8.2)	345 (91.8)
Preferred frequency (n=374)		
Daily	6 (19.4)	53 (15.4)
Weekly	6 (19.4)	131 (38.0)
Monthly	10 (32.3)	147 (42.6)
Never	8 (25.8)	13 (3.8)
Preferred mechanism (n=375)		
Phone calls	4 (12.9)	13 (3.8)
Text messages	5 (16.2)	70 (20.3)
App	12 (38.7)	186 (53.9)
Email	9 (29)	76 (22)

Correlates of mHealth acceptance

The mean score for mHealth acceptance was 4.1 (SD 1.5), with α=.875. In the multivariable model, being a university graduate (β=0.456, *P*=.003), being in a relationship with a partner (β=0.322, *P*=.04), lifetime engagement in sexualized drug use (β=0.489, *P*=.01), and recent engagement in receptive anal sex (β=0.498, *P*=.006) were associated with higher willingness to use mHealth strategies (Table 5).

Table 5. Univariate and multivariable linear regression correlates of mobile health acceptance among participants (n=376).

Variables	Univariate			Multivariable		
	Beta	SE	*P* value	Beta	SE	*P* value
Age (years)	–0.008	0.012	.52	—[a]	—	—
Ethnicity (Malaya)	0.145	0.160	.37	—	—	—
University graduate[b]	0.465	0.158	.003	0.456	0.154	.003
Relationship status (partner)	0.333	0.158	.04	0.322	0.154	.04
Monthly income	–0.001	0.001	.70	—	—	—
Ever had HIV test	–0.003	0.174	.99	—	—	—
Previously diagnosed with STI[c]	0.039	0.117	.82	—	—	—
Ever used pre-exposure prophylaxis	0.164	0.311	.60	—	—	—
Ever used postexposure prophylaxis	0.034	0.306	.91	—	—	—
Experienced childhood physical abuse	0.163	0.158	.30	—	—	—
Experienced childhood sexual assault	–0.017	0.163	.92	—	—	—
Depressive symptoms	0.253	0.160	.12	—	—	—
Ever injected drugs	0.678	0.379	.07	0.408	0.389	.29
Engaged in anal sex (past 6 months)	0.599	0.432	.17	—	—	—
Type of anal sex (past 6 months)						
Insertive	–0.036	0.176	.84	—	—	—
Receptive	0.502	0.183	.006	0.498	0.179	.006
HIV-serodiscordant relationship (past 6 months)	–0.314	0.268	.24	—	—	—
Consistent condom use (past 6 months)	0.130	0.201	.52	—	—	—
Ever engaged in sexualized drug use	0.537	0.189	.005	0.489	0.195	.01

[a]No data because the variable was not included in the model.

[b]This category included college, university, or professional degrees.

[c]STI: sexually transmitted infection.

Discussion

Principal Findings

The rapid growth and use of web-based communication technologies have led to innovations in public health programming and patient care [28], particularly as COVID-19–related restrictions affect health care delivery, with decreased access to in-person health care and prevention interventions, which has had negative consequences on health [43-47]. The utilization of innovative tools in virtual spaces (eHealth or mHealth) can help bridge gaps in service delivery that are needed to improve access to needed health and prevention services, particularly among underserved populations [26]. Until now, there has been a lack of empirical evidence on how communication technologies can improve access to and engagement in care or support for the use of mHealth strategies for HIV prevention needs among men who have sex with men in Malaysia. The findings from this study allow us to assess opportunities to implement mHealth strategies for HIV prevention efforts and inform the specific format and features of a mHealth platform tailored to the needs of men who have sex with men in Malaysia. We found near-universal access to communication technology and internet use and high levels of acceptability of mHealth, particularly smartphone apps, for HIV prevention efforts. Our findings provide preliminary evidence that supports the feasibility of mHealth deployment to deliver HIV prevention and related interventions among men who have sex with men in Malaysia.

We identified a number of interventions that can be incorporated into HIV-prevention mHealth, including tracking and monitoring the sexual and drug use behaviors and assessing symptoms associated with depression. Consistent with information in existing literature [48-50], a considerable proportion of men who have sex with men in our sample had high rates of condomless sex, sexualized drug use, and mental health comorbidities and sexually transmitted infections, highlighting the vulnerabilities in men who have sex with men in an evolving and dynamic HIV epidemic. In addition, it was concerning that a proportion of participants had never been tested for HIV (109/376, 29.0%). HIV testing is the first step in engaging individuals in HIV prevention and treatment cascades; frequent screenings are thus of the utmost importance for individuals most at risk. mHealth can play a uniquely important role in assessing risk and reminding and motivating men who have sex with men to be tested regularly [51-55]. As such, innovations in mHealth can accelerate engagement in HIV testing and

facilitate the uptake of HIV prevention services, such as pre-exposure prophylaxis and HIV self-testing, particularly in areas where such services are underutilized or have limited availability [53]. Our findings underscore the urgent need for innovative strategies to reach and deliver HIV prevention services to this key population, particularly in a context where same-sex sexual behavior is deeply stigmatized.

Near-universal access to smartphones, combined with daily internet access among our sample, consistent with previous studies [56] with Malaysian men who have sex with men, support the feasibility of developing mHealth strategies for Malaysian men who have sex with men. Participants reported already using their smartphones to search for health-related information or apps, and most participants used mobile technologies (ie, smartphones) far more frequently than they used other technologies (eg, computers, landline telephone). This likely reflects the digital revolution that is especially explosive in Asia and the fast-paced growth and use of mobile technologies in the community [57-61]. The rapid advances in mobile technologies and the development of apps open new opportunities for integrating mobile health into existing HIV prevention service delivery in the region. Future research, however, is needed to gather additional information on Malaysian men who have sex with men's interests in health-related content using smartphones, access points (eg, websites, chatbots, apps), which types of information, and where gaps in existing web-based resources exist.

Our findings indicate that there is considerable interest in specific mHealth strategies, such as receiving information related to HIV (345/376, 91.8%), receiving medication reminders (336/376, 89.4%), screening and monitoring sex activity (306/376, 81.4%) and illicit drug use (281/376, 74.7%), and monitoring drug cravings (280/376, 74.5%). In the absence of public dialogue about these issues in Malaysia, it is not surprising that Malaysian men who have sex with men, similar to men who have sex with men in other settings [62-67], and specifically, for HIV prevention, in China [68], Vietnam [69], and Indonesia [66], expressed interest. In our sample, men expressed interest in the use of mHealth apps to receive daily reminders to take medications; however, weekly and monthly reminders were preferred for other activities, with smartphone apps being the preferred platform. An mHealth-based app may serve as an additional tool that can help support men who have sex with men with HIV prevention or care needs between clinical visits, guide them to needed services, enhance clinical care and support through screening and recommendations, and provide different modes of accessing information, services, and prevention commodities. Men who have sex with men at higher risk for HIV due to sexual practices, such as recent engagement in receptive anal sex and lifetime engagement in sexualized drug use, were more willing to use mHealth-based strategies, which suggests that mHealth approaches are particularly promising to address the needs of the subset of men who have sex with men who would most benefit.

While this study focused on HIV prevention behavior, knowledge, and access to services among HIV-negative Malaysian men who have sex with men, mHealth-based health intervention strategies can potentially also play an important role among HIV-positive men who have sex with men, for instance, by encouraging adherence to antiretroviral therapy and increasing retention in the HIV-care cascade [70]. High rates of mHealth acceptability have been reported in HIV-positive men who have sex with men and transgender women in Malaysia, for receiving HIV-related and sexual health information, assessing sexual and health behaviors, and receiving reminders to take HIV medications [71].

Limitations

Despite the many new findings, this study is not without limitations. First, we used self-reported measures for HIV and sexually transmitted infection diagnoses, sexual behavior, and drug use, which may have introduced some social desirability bias and underreporting of stigmatized behaviors. We reduced the potential for social desirability bias by designing the questionnaire to be self-administered and anonymous and by allowing participants to complete the questionnaire alone (without the presence of research staff). Second, though participants were recruited using respondent-driven sampling, some data suggest that a representative sample of all men who have sex with men in Malaysia may not have been achieved. For example, over half of our participants were university graduates, compared with only one-sixth among the general adult population in Malaysia [72]. Men who have sex with men in Malaysia's capital, however, are typically more educated, which perhaps makes the sample more representative of urban men who have sex with men in Malaysia; our findings on participants' educational status are consistent with previous studies conducted among men who have sex with men in the region [49,71]. The educational level of our sample may explain the high level of interest in mHealth and the acceptability of mHealth interventions, as a higher level of education likely facilitates greater technology literacy. This is supported by the multivariable analysis of associations between mHealth acceptance and participant characteristics, which showed that university graduates were more willing to use mHealth strategies. Finally, although men in this study showed a strong willingness to use mHealth for various needs, it should be noted that willingness or interest may not fully reflect actual use; therefore, studies of mHealth interventions in practice are needed to assess use.

To the best of our knowledge, this is the first study to assess the acceptability of mHealth to address HIV prevention needs among men who have sex with men in Malaysia. Our findings show that mHealth use, particularly app-based platforms, appears to be highly acceptable to this population. This finding is particularly meaningful in Malaysia, where there are limited physical venues that are culturally acceptable for men who have sex with men to seek health care since same-sex sexual behavior is illegal, and men who have sex with men are highly stigmatized and are frequent targets of discrimination [9].

Future Implications

Our findings support the development of mHealth-based strategies, especially smartphone apps, to jumpstart the HIV prevention cascade by promoting HIV testing and, depending on the results, linking individuals to the appropriate prevention or treatment services. mHealth strategies, such as culturally

tailored apps, are uniquely positioned to deliver multicomponent interventions, and thus, can bridge systematic gaps needed to address syndemic and complex interrelated health needs and more effective utilization of health services in this underserved population [53,55,73-76]

In recent years, several apps for HIV prevention and treatment efforts have been evaluated in pilot studies or randomized trials that incorporate components such as HIV testing, condom use, pre-exposure prophylaxis, treatment as prevention, and other support services (eg, mental health, drug use) [53,55,73-76]. Some of these apps offer the opportunity to assess, with ecological momentary assessment, or intervene, via ecological momentary interventions, individuals in their natural environment, thereby enabling a better understanding of the factors triggering problems and addressing the problems when and where they arise [77-79].

Most, if not all, of the available apps, however, are primarily developed for use in high-income countries [53,55,73-76]. Additional research to assess the design, functionality, and content preferences of Malaysian men who have sex with men is now needed to facilitate the design of a customized mHealth app in the Malaysian context (specifically, addressing the multiethnic population of Malaysia, as well as men who have sex with men in nonurban settings, will be important). Further research is also needed to understand the perspective of those tasked with providing care and support services via the mHealth platforms (eg, physicians, pharmacists, counselors, outreach workers). Such research will inform and facilitate the integration of mHealth platforms within existing health care services.

Acknowledgments

RS was supported in part by a career development award from the National Institute on Drug Abuse (K01 DA051346, principal investigator: RS) and a research grant from the Fogarty International Center (R21 TW011665; principal investigator: RS). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflicts of Interest

None declared.

References

1. Shao Y, Williamson C. The HIV-1 epidemic: low- to middle-income countries. Cold Spring Harb Perspect Med 2012 Mar 22;2(3):a007187 [FREE Full text] [doi: 10.1101/cshperspect.a007187] [Medline: 22393534]
2. Patel P, Borkowf C, Brooks J, Lasry A, Lansky A, Mermin J. Estimating per-act HIV transmission risk: a systematic review. AIDS 2014 Jun 19;28(10):1509-1519 [FREE Full text] [doi: 10.1097/QAD.0000000000000298] [Medline: 24809629]
3. Sexually transmitted infections treatment guidelines. Center for Disease Control and Prevention. 2021 Jul. URL: https://www.cdc.gov/std/treatment-guidelines/toc.htm [accessed 2022-06-13]
4. Koblin B, Husnik M, Colfax G, Huang Y, Madison M, Mayer K, et al. Risk factors for HIV infection among men who have sex with men. AIDS 2006 Mar 21;20(5):731-739. [doi: 10.1097/01.aids.0000216374.61442.55] [Medline: 16514304]
5. HIV infection risk, prevention, and testing behaviors among men who have sex with men national HIV behavioral surveillance, 23 U.S. cities. Centers for Disease Control and Prevention. 2017. URL: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-special-report-number-22.pdf [accessed 2022-06-13]
6. Global AIDS monitoring 2020: Malaysia HIV/AIDS progress report. Ministry of Health Malaysia. 2019. URL: https://www.moh.gov.my/moh/resources/Penerbitan/Laporan/Umum/Laporan_Global_AIDS_Monitoring_2020_new.pdf [accessed 2022-06-13]
7. Malaysia. Joint United Nations Programme on HIV/AIDS. URL: https://www.unaids.org/en/regionscountries/countries/malaysia [accessed 2022-06-13]
8. Lim SH, Alias H, Kien JKW, Akbar M, Kamarulzaman A, Wong LP. A qualitative study of HIV "test-and-treat" experience among men who have sex with men in Malaysia. AIDS Educ Prev 2019 Jun;31(3):193-205 [FREE Full text] [doi: 10.1521/aeap.2019.31.3.193] [Medline: 31145000]
9. Arreola S, Santos G, Beck J, Sundararaj M, Wilson PA, Hebert P, et al. Sexual stigma, criminalization, investment, and access to HIV services among men who have sex with men worldwide. AIDS Behav 2015 Feb 3;19(2):227-234. [doi: 10.1007/s10461-014-0869-x] [Medline: 25086670]
10. Earnshaw VA, Jin H, Wickersham J, Kamarulzaman A, John J, Altice FL. Exploring intentions to discriminate against patients living with HIV/AIDS among future healthcare providers in Malaysia. Trop Med Int Health 2014 Jun 26;19(6):672-679 [FREE Full text] [doi: 10.1111/tmi.12306] [Medline: 24666546]
11. Jin H, Earnshaw VA, Wickersham JA, Kamarulzaman A, Desai MM, John J, et al. An assessment of health-care students' attitudes toward patients with or at high risk for HIV: implications for education and cultural competency. AIDS Care 2014;26(10):1223-1228 [FREE Full text] [doi: 10.1080/09540121.2014.894616] [Medline: 24625279]
12. Tee YC, Earnshaw VA, Altice FL, Jin H, Kamarulzaman A, Wickersham JA. Evaluating physicians' intention to discriminate against patients living with HIV in Malaysia. AIDS Behav 2019 Apr 17;23(4):1039-1047 [FREE Full text] [doi: 10.1007/s10461-018-2362-4] [Medline: 30560483]

13. Earnshaw VA, Jin H, Wickersham JA, Kamarulzaman A, John J, Lim SH, et al. Stigma toward men who have sex with men among future healthcare providers in Malaysia: would more interpersonal contact reduce prejudice? AIDS Behav 2016 Jan 1;20(1):98-106 [FREE Full text] [doi: 10.1007/s10461-015-1168-x] [Medline: 26324078]
14. Guadamuz TE, Cheung DH, Wei C, Koe S, Lim SH. Young, online and in the dark: scaling up HIV testing among MSM in ASEAN. PLoS One 2015 May 14;10(5):e0126658 [FREE Full text] [doi: 10.1371/journal.pone.0126658] [Medline: 25973907]
15. Lim SH, Bazazi AR, Sim C, Choo M, Altice FL, Kamarulzaman A. High rates of unprotected anal intercourse with regular and casual partners and associated risk factors in a sample of ethnic Malay men who have sex with men (MSM) in Penang, Malaysia. Sex Transm Infect 2013 Dec 20;89(8):642-649. [doi: 10.1136/sextrans-2012-050995] [Medline: 23787168]
16. Chong SCS, Kamarulzaman A, Azwa I, Ng R, Chong M, Raman N, et al. Delayed HIV testing and treatment seeking, and associated support needs among people living with HIV in Malaysia: a qualitative study. Sex Health 2021 May;18(2):147-155. [doi: 10.1071/SH20180] [Medline: 33715762]
17. Chawarski MC, Vicknasingam B, Mazlan M, Schottenfeld RS. Lifetime ATS use and increased HIV risk among not-in-treatment opiate injectors in Malaysia. Drug Alcohol Depend 2012 Jul 01;124(1-2):177-180 [FREE Full text] [doi: 10.1016/j.drugalcdep.2011.12.024] [Medline: 22266088]
18. Fendrich M, Avci O, Johnson TP, Mackesy-Amiti ME. Depression, substance use and HIV risk in a probability sample of men who have sex with men. Addict Behav 2013 Mar;38(3):1715-1718 [FREE Full text] [doi: 10.1016/j.addbeh.2012.09.005] [Medline: 23254224]
19. Daskalopoulou M, Rodger A, Phillips AN, Sherr L, Speakman A, Collins S, et al. Recreational drug use, polydrug use, and sexual behaviour in HIV-diagnosed men who have sex with men in the UK: results from the cross-sectional ASTRA study. Lancet HIV 2014 Oct;1(1):e22-e31 [FREE Full text] [doi: 10.1016/S2352-3018(14)70001-3] [Medline: 26423813]
20. Hirshfield S, Remien RH, Humberstone M, Walavalkar I, Chiasson MA. Substance use and high-risk sex among men who have sex with men: a national online study in the USA. AIDS Care 2004 Nov 27;16(8):1036-1047. [doi: 10.1080/09540120412331292525] [Medline: 15511735]
21. Van Tieu H, Koblin BA. HIV, alcohol, and noninjection drug use. Curr Opin HIV AIDS 2009 Jul;4(4):314-318. [doi: 10.1097/COH.0b013e32832aa902] [Medline: 19532070]
22. Halkitis PN, Parsons JT. Recreational drug use and HIV-risk sexual behavior among men frequenting gay social venues. J Gay Lesbian Soc Serv 2008 Sep 22;14(4):19-38. [doi: 10.1300/J041v14n04_02]
23. Sandfort T, Knox J, Alcala C, El-Bassel N, Kuo I, Smith LR. Substance use and HIV risk among men who have sex with men in africa: a systematic review. J Acquir Immune Defic Syndr 2017;76(2):e34-e46. [doi: 10.1097/qai.0000000000001462]
24. Cooper V, Clatworthy J, Whetham J, Consortium E. mHealth interventions to support self-management in HIV: a systematic review. Open AIDS J 2017 Nov 21;11(1):119-132 [FREE Full text] [doi: 10.2174/1874613601711010119] [Medline: 29290888]
25. Henny KD, Wilkes AL, McDonald CM, Denson DJ, Neumann MS. A rapid review of ehealth interventions addressing the continuum of HIV care (2007-2017). AIDS Behav 2018 Jan 5;22(1):43-63 [FREE Full text] [doi: 10.1007/s10461-017-1923-2] [Medline: 28983684]
26. Conserve DF, Jennings L, Aguiar C, Shin G, Handler L, Maman S. Systematic review of mobile health behavioural interventions to improve uptake of HIV testing for vulnerable and key populations. J Telemed Telecare 2017 Feb;23(2):347-359 [FREE Full text] [doi: 10.1177/1357633X16639186] [Medline: 27056905]
27. 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]
28. mHealth: new horizons for health through mobile technologies: second global survey on eHealth. World Health Organization. 2011. URL: https://apps.who.int/iris/handle/10665/44607 [accessed 2022-06-13]
29. Free C, Phillips G, Watson L, Galli L, Felix L, Edwards P, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med 2013 Jan 15;10(1):e1001363 [FREE Full text] [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
30. Lester RT, Ritvo P, Mills EJ, Kariri A, Karanja S, Chung MH, et al. Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomised trial. Lancet 2010 Nov 27;376(9755):1838-1845. [doi: 10.1016/S0140-6736(10)61997-6] [Medline: 21071074]
31. Milward J, Lynskey M, Strang J. Solving the problem of non-attendance in substance abuse services. Drug Alcohol Rev 2014 Nov 06;33(6):625-636. [doi: 10.1111/dar.12194] [Medline: 25196817]
32. Mbuagbaw L, Mursleen S, Lytvyn L, Smieja M, Dolovich L, Thabane L. Mobile phone text messaging interventions for HIV and other chronic diseases: an overview of systematic reviews and framework for evidence transfer. BMC Health Serv Res 2015 Jan 22;15(1):33-16 [FREE Full text] [doi: 10.1186/s12913-014-0654-6] [Medline: 25609559]
33. Hand phone users survey 2018. Malaysian Communications and Multimedia Commission. 2019. URL: https://www.mcmc.gov.my/skmmgovmy/media/General/pdf/HPUS2018.pdf [accessed 2022-06-13]
34. Heckathorn DD. Respondent-driven sampling: a new approach to the study of hidden populations. Soc Probl 1997 May;44(2):174-199. [doi: 10.1525/sp.1997.44.2.03x0221m]

35. Krishnan A, Ferro EG, Weikum D, Vagenas P, Lama JR, Sanchez J, et al. Communication technology use and mHealth acceptance among HIV-infected men who have sex with men in Peru: implications for HIV prevention and treatment. AIDS Care 2015 Oct 06;27(3):273-282 [FREE Full text] [doi: 10.1080/09540121.2014.963014] [Medline: 25285464]
36. Shrestha R, Karki P, Copenhaver M. Interest in use of mHealth technology in HIV prevention and associated factors among high-risk drug users enrolled in methadone maintenance program. AIDS Care 2017 Sep 08;29(9):1144-1148 [FREE Full text] [doi: 10.1080/09540121.2017.1325439] [Medline: 28478706]
37. Shrestha R, Huedo-Medina TB, Altice FL, Krishnan A, Copenhaver M. Examining the acceptability of mhealth technology in HIV prevention among high-risk drug users in treatment. AIDS Behav 2017 Nov 26;21(11):3100-3110 [FREE Full text] [doi: 10.1007/s10461-016-1637-x] [Medline: 28025735]
38. Behavioral risk factor surveillance system (BRFSS) questionnaire. Centers for Disease Control and Prevention. 2006. URL: https://www.cdc.gov/brfss/index.html [accessed 2022-06-13]
39. Radloff LS. The CES-D scale. Appl Psychol Meas 2016 Jul 26;1(3):385-401 [FREE Full text] [doi: 10.1177/014662167700100306]
40. Björgvinsson T, Kertz SJ, Bigda-Peyton JS, McCoy KL, Aderka IM. Psychometric properties of the CES-D-10 in a psychiatric sample. Assessment 2013 Aug 18;20(4):429-436. [doi: 10.1177/1073191113481998] [Medline: 23513010]
41. Sogawa C, Sogawa N, Tagawa J, Fujino A, Ohyama K, Asanuma M, et al. 5-Methoxy-N,N-diisopropyltryptamine (Foxy), a selective and high affinity inhibitor of serotonin transporter. Toxicol Lett 2007 Apr 05;170(1):75-82. [doi: 10.1016/j.toxlet.2007.02.007] [Medline: 17382495]
42. US Preventive Services Task Force, Owens DK, Davidson KW, Krist AH, Barry MJ, Cabana M, et al. Preexposure prophylaxis for the prevention of HIV infection: US Preventive Services Task Force recommendation statement. JAMA 2019 Jun 11;321(22):2203-2213. [doi: 10.1001/jama.2019.6390] [Medline: 31184747]
43. Cantor J, Sood N, Bravata DM, Pera M, Whaley C. The impact of the COVID-19 pandemic and policy response on health care utilization: evidence from county-level medical claims and cellphone data. J Health Econ 2022 Mar;82:102581-102516 [FREE Full text] [doi: 10.1016/j.jhealeco.2022.102581] [Medline: 35067386]
44. Malhotra C, Chaudhry I, Ozdemir S, Finkelstein EA. Reduced health-care utilization among people with chronic medical conditions during coronavirus disease 2019. Proceedings of Singapore Healthcare 2020 Oct 07;30(3):254-257. [doi: 10.1177/2010105820964533]
45. Sanchez TH, Zlotorzynska M, Rai M, Baral SD. Characterizing the impact of COVID-19 on men who have sex with men across the United States in April, 2020. AIDS Behav 2020 Jul 29;24(7):2024-2032 [FREE Full text] [doi: 10.1007/s10461-020-02894-2] [Medline: 32350773]
46. Santos G, Ackerman B, Rao A, Wallach S, Ayala G, Lamontage E, et al. Economic, mental health, HIV prevention and HIV treatment impacts of COVID-19 and the COVID-19 response on a global sample of cisgender gay men and other men who have sex with menmental health, HIV prevention and HIV treatment impacts of COVID-19 and the COVID-19 response on a global sample of cisgender gay men and other men who have sex with men. AIDS Behav 2021 Feb 11;25(2):311-321 [FREE Full text] [doi: 10.1007/s10461-020-02969-0] [Medline: 32654021]
47. Mistler CB, Curley CM, Rosen AO, El-Krab R, Wickersham JA, Copenhaver MM, et al. The impact of COVID-19 on access to HIV prevention services among opioid-dependent individuals. J Community Health 2021 Oct 26;46(5):960-966 [FREE Full text] [doi: 10.1007/s10900-021-00979-0] [Medline: 33770334]
48. Lim SH, Mburu G, Bourne A, Pang J, Wickersham JA, Wei CKT, et al. Willingness to use pre-exposure prophylaxis for HIV prevention among men who have sex with men in Malaysia: Findings from an online survey. PLoS One 2017 Sep 13;12(9):e0182838 [FREE Full text] [doi: 10.1371/journal.pone.0182838] [Medline: 28902857]
49. Kanter J, Koh C, Razali K, Tai R, Izenberg J, Rajan L, et al. Risk behaviour and HIV prevalence among men who have sex with men in a multiethnic society: a venue-based study in Kuala Lumpur, Malaysia. Int J STD AIDS 2011 Jan 01;22(1):30-37. [doi: 10.1258/ijsa.2010.010277] [Medline: 21364064]
50. Shrestha R, Alias H, Wong LP, Altice FL, Lim SH. Using individual stated-preferences to optimize HIV self-testing service delivery among men who have sex with men (MSM) in Malaysia: results from a conjoint-based analysis. BMC Public Health 2020 Nov 25;20(1):1777-1711 [FREE Full text] [doi: 10.1186/s12889-020-09832-w] [Medline: 33238941]
51. Zhu X, Zhang W, Operario D, Zhao Y, Shi A, Zhang Z, et al. Effects of a mobile health intervention to promote HIV self-testing with MSM in China: a randomized controlled trial. AIDS Behav 2019 Nov 9;23(11):3129-3139 [FREE Full text] [doi: 10.1007/s10461-019-02452-5] [Medline: 30852728]
52. Sanabria G, Scherr T, Garofalo R, Kuhns LM, Bushover B, Nash N, et al. Usability evaluation of the mLab app for improving home HIV testing behaviors in youth at risk of HIV infection. AIDS Educ Prev 2021 Aug;33(4):312-324 [FREE Full text] [doi: 10.1521/aeap.2021.33.4.312] [Medline: 34370566]
53. Biello KB, Hill-Rorie J, Valente PK, Futterman D, Sullivan PS, Hightow-Weidman L, et al. Development and evaluation of a mobile app designed to increase HIV testing and pre-exposure prophylaxis use among young men who have sex with men in the United States: open pilot trial. J Med Internet Res 2021 Mar 24;23(3):e25107 [FREE Full text] [doi: 10.2196/25107] [Medline: 33759792]

54. Biello KB, Horvitz C, Mullin S, Mayer KH, Scott H, Coleman K, et al. HIV self-testing and STI self-collection via mobile apps: experiences from two pilot randomized controlled trials of young men who have sex with men. Mhealth 2021 Apr;7:26-26 [FREE Full text] [doi: 10.21037/mhealth-20-70] [Medline: 33898595]
55. Sullivan PS, Hightow-Weidman L. Mobile apps for HIV prevention: how do they contribute to our epidemic response for adolescents and young adults? Mhealth 2021 Apr;7:36-36 [FREE Full text] [doi: 10.21037/mhealth-20-71] [Medline: 33898605]
56. Shrestha R, Lim SH, Altice FL, Copenhaver M, Wickersham JA, Saifi R, et al. Use of smartphone to seek sexual health information online among Malaysian men who have sex with men (MSM): implications for mhealth intervention to increase HIV testing and reduce HIV risks. J Community Health 2020 Feb 2;45(1):10-19 [FREE Full text] [doi: 10.1007/s10900-019-00713-x] [Medline: 31375976]
57. Sedik TS, Chen S, Feyzioglu T, Ghazanchyan M, Gupta S, Jahan S, et al. The digital revolution in Asia and its macroeconomic effects. ADBI Working Paper Series. 2019. URL: https://www.adb.org/sites/default/files/publication/535846/adbi-wp1029.pdf [accessed 2022-06-26]
58. Li B, Piachaud D. Technological innovations and social development in Asia. J Asian Public Policy 2018 Nov 23;12(1):1-14. [doi: 10.1080/17516234.2018.1546419]
59. Hossain M, Weng W, Bhattacharya S, Majumder H, Faizah F. Digital health divide in South Asia: ethical concerns, challenges, and recommendations. SocArxiv Preprint posted online on April 1, 2020. [doi: 10.31235/osf.io/qxecj]
60. Baur A, Yew H, Xin M. The future of healthcare in Asia: digital health ecosystems. McKinsey & Company. 2021. URL: https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/the-future-of-healthcare-in-asia-digital-health-ecosystems [accessed 2022-06-13]
61. Seman S. Are we ready to app?: a study on mHealth apps, its future, and trends in Malaysia context. In: Ramayah T, editor. Mobile Platforms, Design, and Apps for Social Commerce. Hershey, Pennsylvania, USA: IG Global; 2017:1-15.
62. Lu X, Gao P, Wang X, Wang X, Zhao X, He Q, et al. User preferences for an mhealth approach to support HIV self-testing and linkage to HIV prevention or care services for MSM in China. AIDS Educ Prev 2021 Dec;33(6):534-550. [doi: 10.1521/aeap.2021.33.6.534] [Medline: 34874759]
63. Aliabadi N, Carballo-Dieguez A, Bakken S, Rojas M, Brown W, Carry M, et al. Using the information-motivation-behavioral skills model to guide the development of an HIV prevention smartphone application for high-risk MSM. AIDS Educ Prev 2015 Dec;27(6):522-537 [FREE Full text] [doi: 10.1521/aeap.2015.27.6.522] [Medline: 26595265]
64. Goldenberg T, McDougal SJ, Sullivan PS, Stekler JD, Stephenson R. Preferences for a mobile HIV prevention app for men who have sex with men. JMIR Mhealth Uhealth 2014 Oct 29;2(4):e47 [FREE Full text] [doi: 10.2196/mhealth.3745] [Medline: 25355249]
65. Zhao Y, Zhu X, Pérez AE, Zhang W, Shi A, Zhang Z, et al. MHealth approach to promote oral HIV self-testing among men who have sex with men in China: a qualitative description. BMC Public Health 2018 Sep 27;18(1):1146-1148 [FREE Full text] [doi: 10.1186/s12889-018-6046-9] [Medline: 30261856]
66. Garg PR, Uppal L, Mehra S, Mehra D. Mobile health app for self-learning on HIV prevention knowledge and services among a young Indonesian key population: cohort study. JMIR Mhealth Uhealth 2020 Sep 08;8(9):e17646 [FREE Full text] [doi: 10.2196/17646] [Medline: 32896831]
67. Holloway IW, Winder TJ, Lea CH, Tan D, Boyd D, Novak D. Technology use and preferences for mobile phone–based HIV prevention and treatment among black young men who have sex with men: exploratory research. JMIR Mhealth Uhealth 2017 Apr 13;5(4):e46 [FREE Full text] [doi: 10.2196/mhealth.6436] [Medline: 28408360]
68. Marley G, Fu G, Zhang Y, Li J, Tucker JD, Tang W, et al. Willingness of Chinese men who have sex with men to use smartphone-based electronic readers for HIV self-testing: web-based cross-sectional study. J Med Internet Res 2021 Nov 19;23(11):e26480 [FREE Full text] [doi: 10.2196/26480] [Medline: 34806988]
69. Trang K, Le LX, Brown CA, To MQ, Sullivan PS, Jovanovic T, et al. Feasibility, acceptability, and design of a mobile ecological momentary assessment for high-risk men who have sex with men in Hanoi, Vietnam: qualitative study. JMIR Form Res 2022 Jan 27;6(1):e30360 [FREE Full text] [doi: 10.2196/30360] [Medline: 35084340]
70. Muessig K, LeGrand S, Horvath K, Bauermeister J, Hightow-Weidman LB. Recent mobile health interventions to support medication adherence among HIV-positive MSM. Curr Opin HIV AIDS 2017 Sep;12(5):432-441 [FREE Full text] [doi: 10.1097/COH.0000000000000401] [Medline: 28639990]
71. Krishnan A, Weikum D, Cravero C, Kamarulzaman A, Altice FL. Assessing mobile technology use and mHealth acceptance among HIV-positive men who have sex with men and transgender women in Malaysia. PLoS One 2021 Mar 23;16(3):e0248705 [FREE Full text] [doi: 10.1371/journal.pone.0248705] [Medline: 33755693]
72. Graduates statistics 2020. Department of Statistics Malaysia. 2021. URL: https://www.dosm.gov.my/v1/index.php?r=column/cthemeByCat&cat=476&bul_id=U1ltVWpwNXRNRUR2NlhRSHZmenRMUT09&menu_id=Tm8zcnRjdVRNWWlpWjRlbmtlaDk1UT09 [accessed 2022-06-13]
73. Sullivan PS, Driggers R, Stekler JD, Siegler A, Goldenberg T, McDougal SJ, et al. Usability and acceptability of a mobile comprehensive HIV prevention app for men who have sex with men: a pilot study. JMIR Mhealth Uhealth 2017 Mar 09;5(3):e26 [FREE Full text] [doi: 10.2196/mhealth.7199] [Medline: 28279949]

74. Canan CE, Waselewski ME, Waldman ALD, Reynolds G, Flickinger TE, Cohn WF, et al. Long term impact of PositiveLinks: Clinic-deployed mobile technology to improve engagement with HIV care. PLoS One 2020;15(1):e0226870 [FREE Full text] [doi: 10.1371/journal.pone.0226870] [Medline: 31905209]
75. Hung R, Bandy J, Wantuck J. PlushCare tele-PrEP, a nationwide app-based program's California success. 2019 Presented at: National HIV Prevention Conference; March 18-21, 2019; Altlanta, Georgia.
76. Wong KYK, Stafylis C, Klausner JD. Telemedicine: a solution to disparities in human immunodeficiency virus prevention and pre-exposure prophylaxis uptake, and a framework to scalability and equity. Mhealth 2020;6:21 [FREE Full text] [doi: 10.21037/mhealth.2019.12.06] [Medline: 32270013]
77. Fatseas M, Serre F, Swendsen J, Auriacombe M. Effects of anxiety and mood disorders on craving and substance use among patients with substance use disorder: an ecological momentary assessment study. Drug Alcohol Depend 2018 Jun 01;187:242-248. [doi: 10.1016/j.drugalcdep.2018.03.008] [Medline: 29684892]
78. Jones A, Tiplady B, Houben K, Nederkoorn C, Field M. Do daily fluctuations in inhibitory control predict alcohol consumption? An ecological momentary assessment study. Psychopharmacology (Berl) 2018 May 1;235(5):1487-1496 [FREE Full text] [doi: 10.1007/s00213-018-4860-5] [Medline: 29497782]
79. Merikangas KR, Swendsen J, Hickie IB, Cui L, Shou H, Merikangas AK, et al. Real-time mobile monitoring of the dynamic associations among motor activity, energy, mood, and sleep in adults with bipolar disorder. JAMA Psychiatry 2019 Feb 01;76(2):190-198 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.3546] [Medline: 30540352]

Abbreviations

mHealth: mobile health
MYR: Malaysian Ringgit

Edited by A Mavragani; submitted 29.01.22; peer-reviewed by D Brennan, B Wong, N Xiang, Y Nee, K Gautam; comments to author 11.03.22; revised version received 14.04.22; accepted 08.05.22; published 25.07.22.

Please cite as:
Shrestha R, Maviglia F, Altice FL, DiDomizio E, Khati A, Mistler C, Azwa I, Kamarulzaman A, Halim MAA, Wickersham JA Mobile Health Technology Use and the Acceptability of an mHealth Platform for HIV Prevention Among Men Who Have Sex With Men in Malaysia: Cross-sectional Respondent-Driven Sampling Survey
J Med Internet Res 2022;24(7):e36917
URL: https://www.jmir.org/2022/7/e36917
doi:10.2196/36917
PMID:35877172

Original Paper

Initial Psychometric Properties of 7 NeuroUX Remote Ecological Momentary Cognitive Tests Among People With Bipolar Disorder: Validation Study

Raeanne C Moore[1], PhD; Emma M Parrish[1,2], MS; Ryan Van Patten[3,4], PhD; Emily Paolillo[1,2], PhD; Tess F Filip[1], BA; Jessica Bomyea[1,5], PhD; Derek Lomas[6], PhD; Elizabeth W Twamley[1,5], PhD; Lisa T Eyler[1,5], PhD; Colin A Depp[1,5], PhD

[1]Department of Psychiatry, University of California San Diego, San Diego, CA, United States
[2]SDSU/UC San Diego Joint Doctoral Program in Clinical Psychology, San Diego, CA, United States
[3]Psychiatry and Human Behavior, Brown University, Providence, RI, United States
[4]Providence VA Medical Center, Providence, RI, United States
[5]VA San Diego Healthcare System, San Diego, CA, United States
[6]Industrial Design Engineering, Delft University of Technology, Delft, Netherlands

Corresponding Author:
Raeanne C Moore, PhD
Department of Psychiatry
University of California San Diego
220 Dickinson St. Ste B (8231)
San Diego, CA, 92103-8231
United States
Phone: 1 949 933 8063
Email: r6moore@health.ucsd.edu

Abstract

Background: As smartphone technology has become nearly ubiquitous, there is a growing body of literature suggesting that ecological momentary cognitive testing (EMCT) offers advantages over traditional pen-and-paper psychological assessment. We introduce a newly developed platform for the self-administration of cognitive tests in ecologically valid ways.

Objective: The aim of this study is to develop a Health Insurance Portability and Accountability Act–compliant EMCT smartphone-based platform for the frequent and repeated testing of cognitive abilities in everyday life. This study examines the psychometric properties of 7 mobile cognitive tests covering domains of processing speed, visual working memory, recognition memory, and response inhibition within our platform among persons with and without bipolar disorder (BD). Ultimately, if shown to have adequate psychometric properties, EMCTs may be useful in research on BD and other neurological and psychiatric illnesses.

Methods: A total of 45 persons with BD and 21 demographically comparable healthy volunteer participants (aged 18-65 years) completed smartphone-based EMCTs 3 times daily for 14 days. Each EMCT session lasted approximately 1.5 minutes. Only 2 to 3 tests were administered in any given session, no test was administered more than once per day, and alternate test versions were administered in each session.

Results: The mean adherence to the EMCT protocol was 69.7% (SD 20.5%), resulting in 3965 valid and complete tests across the full sample. Participants were significantly more likely to miss tests on later versus earlier study days. Adherence did not differ by diagnostic status, suggesting that BD does not interfere with EMCT participation. In most tests, age and education were related to EMCT performance in expected directions. The average performances on most EMCTs were moderately to strongly correlated with the National Institutes of Health Toolbox Cognition Battery. Practice effects were observed in 5 tests, with significant differences in practice effects by BD status in 3 tests.

Conclusions: Although additional reliability and validity data are needed, this study provides initial psychometric support for EMCTs in the assessment of cognitive performance in real-world contexts in BD.

(J Med Internet Res 2022;24(7):e36665) doi:10.2196/36665

KEYWORDS

neuropsychology; mobile health; ambulatory assessment; ecological momentary assessment; practice effects; validity; testing; serious mental illness; mobile phone

Introduction

Background

Advances in digital technology and the increasing ubiquity of both internet access [1] and mobile phones [2] are changing the way in which social, emotional, and cognitive states are measured. Ecological momentary assessments (EMAs) [3] deliver brief, repeated, self-reported surveys to examinees as they go about their daily lives. EMAs are designed to capture naturalistic fluctuations in psychological and physiological states in real time without relying on retrospective recall. A growing body of literature supports the utility of EMAs in a variety of populations [4-9], and EMAs have shown greater sensitivity to psychological distress compared with single-time psychological assessments. For example, Moore et al [10] reported that EMA measures were more sensitive to subtle reductions in depression and anxiety after a mindfulness intervention than identical paper-pencil instruments measuring the same constructs.

EMA methods are increasingly being used to assess cognitive performance through brief, repeated neuropsychological tasks. In this context, we refer to the concurrent administration of EMA questions and mobile cognitive tests as ecological momentary cognitive testing (EMCT). Owing to the dynamic nature of cognitive processes with influence from a variety of situational factors [11], single-administration assessment data from traditional in-person cognitive testing are vulnerable to the acute impact of confounds such as sleep deprivation and mood instability. In contrast, EMCT data can be aggregated across multiple testing sessions, thereby potentially reducing the influence of intraindividual error from extraneous variables [12-14]. In addition, EMCT data can be used to directly examine intraindividual variability in cognitive performance across time and contexts. Overall, the delivery of EMCT measures to individuals in real-world environments likely improves ecological validity over controlled laboratory or clinic-based evaluations [12,13,15-17]. In contrast, an advantage of traditional neuropsychological testing over EMCT is the precise control over the environment, allowing for optimization of performance and in-depth measurement of cognitive functioning. Accordingly, EMCT is intended to augment rather than replace comprehensive neuropsychological evaluations [14,18].

Bipolar disorder (BD) may be a particularly appropriate condition for which to implement EMCTs given the symptom fluctuation that characterizes the disorder [19]. Moreover, cognitive sequelae of BD are well recognized [20,21], and cognitive performance is more variable in patients with BD than in controls [21]. Importantly, symptom severity and variability in BD do not prevent the collection of EMA data as multiple studies of psychological constructs have been successfully carried out in patients with BD [22-26].

Objectives

The goal of this study was to investigate the psychometric properties of 7 newly developed mobile cognitive tests administered within an EMCT platform (NeuroUX) in adults with BD and healthy comparison participants. These tests were designed to measure processing speed, working memory, visual memory, verbal recognition memory, reasoning, and inhibitory control. We examined adherence, practice and fatigue effects, intraindividual variability, reliability and validity metrics, and performance differences based on BD status.

EMA adherence in BD, depression, and schizophrenia has varied across studies, with most reporting rates of at least 65% and sometimes >90% [24,27-31]. This study included both EMA and mobile cognitive testing (EMCT) as opposed to EMA only, which could have affected the participants' engagement. Meta-analytic studies of compliance with EMA protocols [30,32,33] range from median adherence rates of 75% to 80%. It is important to note that approximately 10% of participants in EMA studies have significantly lower adherence (approximately 45% [30]), which lowers the mean adherence rates. We aimed for a benchmark of 70% mean adherence in our sample given the extra time commitment needed to complete EMA surveys and mobile cognitive tests in each session.

Phone type differences were examined to address device variability, such as differences in response time latencies that have the potential to affect the recorded test performance [34,35]. To examine convergent and discriminant validity, we also compared aggregate mean scores on the mobile cognitive tests with the National Institutes of Health Toolbox Cognition Battery (NIH-TB-CB), iPad version [36,37], and examined intraindividual variability between groups. We hypothesized that (1) there would be adequate adherence to the EMCT protocol and small practice effects across the full sample, and these metrics would not vary by diagnostic status; (2) EMCT scores would be moderately associated with laboratory-based neuropsychological test scores, and these associations would be strongest among clinical tests assessing the same or similar cognitive processes—specifically, that our mobile cognitive tests of processing speed (Matching Pair, Odd One Out [time to complete], and Quick Tap 1) would be related to global cognition (ie, the National Institutes of Health Toolbox [NIH-TB] Fluid Composite) and to all the individual cognitive domains; that our mobile cognitive tests of working memory (Memory Matrix, Odd One Out [total score], and CopyKat) would be related to global cognition and tests in the domains of working memory, processing speed, and executive function; that our recognition memory mobile cognitive test (Mobile Variable Difficulty List Memory Test [VLMT]) would be related to the NIH-TB memory test and, although less strongly, to global cognition; and that our response inhibition mobile cognitive test (Quick Tap 2) would be related to laboratory-based tests of attention processing speed, working memory, and executive function. We also hypothesized that poorer performance on the mobile cognitive tests would be associated with a diagnosis of

BD, older age, and fewer years of education and that intraindividual variability would be higher in the BD group than in the healthy volunteer group.

Methods

Recruitment

Participants were recruited through flyers and web-based recruitment portals and were either individuals with BD (n=45) or demographically comparable healthy volunteers (n=21). The inclusion criteria were as follows: (1) a diagnosis of BD on the MINI International Neuropsychiatric Interview version 6.0.0 [38] or no psychiatric diagnoses for healthy volunteers, (2) outpatient treatment status for the BD group, (3) age between 18 and 65 years, (4) fluency in English, (5) capacity to provide written informed consent, and (6) not being on conservatorship. The exclusion criteria were as follows: (1) history of neurological disorder or head trauma with loss of consciousness for >15 minutes, (2) sensory impairment, (3) substance use disorder in the previous 3 months (excluding cannabis and tobacco), (4) severe manic symptoms as measured by a score >20 on the Young Mania Rating Scale [39] or severe depressive symptoms as measured by a score >30 the Montgomery-Asberg Depression Rating Scale for the BD group [40], (5) ideation score type 3 or higher on the Columbia Suicide Severity Rating Scale (C-SSRS) [41] in the previous month, and (6) concurrent enrollment in another research study.

Ethics Approval

This study was approved by the University of California San Diego Institutional Review Board (protocol #172120), and all participants provided written informed consent and demonstrated capacity to consent based on a brief screening measure [42]. At the screening visit, the participants completed interview-rated symptom measures and the C-SSRS.

Procedure

At the baseline visit, the participants completed self-report questionnaires and laboratory-based neuropsychological performance tests, including the NIH-TB-CB and Delis-Kaplan Executive Function System (D-KEFS) Color-Word Inference Test. The NIH-TB-CB iPad version was used, which includes a mix of self-administered and examiner-assisted–administered tests. An examiner is required to be present for the full battery, which takes approximately 30 to 45 minutes to administer. The D-KEFS Color-Word Inference Test is examiner-administered and takes approximately 3 minutes to administer. The participants could either use their own smartphones or borrow a study-provided Apple iPhone 7 for home-based EMCT. The participants were trained on the EMCT protocol, and all of them (66/66, 100%) completed a laboratory-based session of the 7 mobile cognitive tests with time for technical questions and troubleshooting. Participants also received an EMCT operating manual, which we developed for this study. The manual included information on when to expect the alerts to take the mobile cognitive tests, directions for how to complete the EMA surveys and mobile cognitive tests, important tips and reminders (eg, reminder to charge their smartphone nightly), troubleshooting tips, and frequently asked questions.

The 14-day EMCT protocol began the day following the baseline visit and consisted of 3 SMS text message notifications per day, with each notification including a link to an EMA survey and 2 to 3 mobile cognitive tests (Table 1). The timing of these testing windows was adjusted according to each person's sleep and wake schedule, and there was a 2-hour minimum between each testing session; links were active for 1 hour after the SMS text message notification. Each of the 7 mobile cognitive tests was administered 9 times over the 14-day testing period (with the exception of the Mobile VLMT, which was administered daily), and the order was counterbalanced to ensure that each test was administered evenly across the morning, midday, and early evening. Different versions of the tasks were administered at each time point. No identifying information was affiliated with the EMCT platform, and deidentified data were instantly uploaded to Amazon Web Services (Health Insurance Portability and Accountability Act–compliant).

Table 1. Protocol for mobile cognitive testing administration.

Mobile cognitive test	Study day													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Matching Pair	MA[a]	EE[b]	EE	EE	MA	—[c]	—	—	MD[d]	—	—	MA	MD	MD
Memory Matrix	—	MD	—	MA	MD	MA	EE	—	EE	EE	—	EE	—	MA
Odd One Out	—	MA	—	MD	EE	—	—	EE	MA	MA	MD	MA	—	EE
VLMT[e] recall	EE	MA	MA	EE	MA	MD	MD	MD	EE	MA	EE	MD	MD	EE
VLMT recognition[f]	EE	MA, MD	MA, MD	EE	MA, MD	MD, EE	MD	MD, EE	EE	MA, MD	EE	MD, EE	MD, EE	EE
Quick Tap 1	EE	—	MA	—	—	MD	MD	MD	—	EE	MA	MD	—	MA
Quick Tap 2	EE	—	MA	—	—	MD	MD	MD	—	EE	MA	MD	—	MA
CopyKat	MD	—	MD	—	—	EE	MA	MA	MD	—	EE	EE	MA	—

[a]MA: morning administration.

[b]EE: early evening administration.

[c]Empty cells indicate that a session was not scheduled at that day and time.

[d]MD: midday administration.

[e]VLMT: Variable Difficulty List Memory Test.

[f]The VLMT recognition was sometimes administered at 2 time points on the same day.

The study staff contacted participants by telephone on the first day of the at-home

protocol, as well as if participants missed >3 surveys in a row, to increase adherence and help troubleshoot any problems. Participants were provided with study staff information, who were available Monday to Friday (and often on Saturdays) to respond to questions and help troubleshoot problems. Following the 14-day EMCT data collection period, participants returned to the laboratory to complete the C-SSRS again and a study feedback questionnaire as well as to return the study smartphone (if one was borrowed). Participants were compensated for all study visits and were given a bonus compensation of US $1 for each EMCT session completed.

Measures

Baseline Symptom Measures

Participants with BD were assessed for depression using the interview-rated Montgomery-Asberg Depression Rating Scale, with total scores ranging from 0 to 60 and higher scores indicating greater severity [40]. Symptoms of mania were assessed using the Young Mania Rating Scale, with a total score of 0 to 60 and higher scores indicating greater severity [39].

Baseline Cognitive Measures

The NIH-TB-CB [37] consists of 7 tests designed to measure language (Oral Reading Recognition), episodic memory (Picture Sequence Memory Test), working memory (List Sorting Working Memory Test), attention (Flanker), executive functioning (Flanker and Dimensional Change Card Sort Test), and processing speed (Pattern Comparison Processing Speed Test). Scores from the attention, episodic memory, working memory, executive functioning, and processing speed tests are averaged to create a Fluid Cognition score. The NIH-TB-CB also includes language tests (receptive vocabulary and reading) as an estimate of premorbid functioning and education (Crystallized Intelligence score). The NIH-TB-CB was normed in a large sample representative of the US population aged 3 to 85 years. The age-corrected standard scores were used in this study [43]. We chose not to use demographically adjusted T-scores to conduct more direct comparisons between NIH-TB-CB test scores and our mobile cognitive tests.

EMCT Platform

The development of our EMCT platform was supported in part by the National Institute of Mental Health. The mobile cognitive tests were designed using a human-centered approach and gamified to increase engagement [44]. Test stimuli differed at each administration, thus serving as alternate forms. The platform was first beta-tested in 10 healthy adults across their life span, and modifications were made based on user feedback. A total of 7 mobile cognitive tests were initially developed and tested in this study. Refer to Table 2 for a list of the mobile cognitive tests, the cognitive domains assessed, completion times, and screenshots.

Table 2. Mobile cognitive tests.

Mobile cognitive test	Cognitive domain assessed	Time to complete	Screenshot of task
Matching Pair	Processing speed	90 seconds (fixed time)	
Memory Matrix	Visual working memory	Variable; 3 trials; approximately 1 to 2 minutes (mean completion time 1.5, SD 1.4 minutes)	
Odd One Out	Visual working memory (primary); processing speed (secondary)	Variable; 9 trials; approximately 1 minute (mean completion time 0.75, SD 0.57 minutes)	
VLMT[a]	Recognition memory	30 seconds for list presentation (fixed time)	
Quick Tap 1	Processing speed	Variable; 12 trials; approximately 1 minute (mean completion time 60.8, SD 64.6 seconds)	
Quick Tap 2	Response inhibition	Variable; 12 trials; approximately 1 minute (mean completion time 66.8, SD 20.6 seconds)	
CopyKat	Visual working memory	Variable; 3 trials; approximately 2 to 3 minutes (mean completion time 2.7, SD 4.0 minutes)	

[a]VLMT: Variable Difficulty List Memory Test.

Mobile Cognitive Test Descriptions

Matching Pair

The cognitive domain assessed is processing speed, and the time to complete is 90 seconds. In Matching Pair, participants are presented with a matrix containing tiles of varying colors and shapes. This tile matrix starts as a 2×2 grid (4 tiles) and gradually increases to a maximum grid of 4×4 (16 tiles). Participants select the 2 tiles that match in color as quickly as possible using 1 finger. Scoring is a weighted accuracy score calculated according to the grid size shown. Faster response times are associated with the possibility of higher scores. The grid size is multiplied (eg, 4 × 4 = 16) and added to the running score of the previous correct trial. For example, for a correct trial with a grid size of 3×3 and a previous trial score of 246, the trial score would be calculated as follows: 3 × 3 = 9 and 9 + 246 = a trial score of 255. If a pair is incorrectly selected, the score does not change. The reaction time for each response is also recorded.

Memory Matrix

The cognitive domain assessed is visual working memory, and the time to complete is variable (3 trials for approximately 1 to 2 minutes total; mean completion time 1.5, SD 1.4 minutes). In Memory Matrix, participants are presented with a matrix of tiles starting with 2×2 (4 tiles), which gradually increases to a maximum of 7×7 (49 tiles). A pattern of contrasting color tiles is presented, and participants are asked to learn the location of these highlighted tiles. The contrasting pattern disappears after 1.5 seconds, and participants touch the tiles that were previously presented. This sequence begins with 1 highlighted tile, and based on performance, it can increase to a maximum of 11 tiles. The number of highlighted tiles increases by 1 for each correct response and decreases by 1 for each incorrect response. The task terminates when the participant makes 3 incorrect responses. For each correct response, the number of highlighted tiles is added to the previous correct score. For example, if the previous trial score was 12 and a correct response is given in a trial with 4 highlighted cells, this trial's new score would be calculated as 16. The score does not change if a trial response is incorrect. Reaction time is also recorded.

Odd One Out

The cognitive domain assessed is visual working memory (primary) and processing speed (secondary), and the time to complete is variable (9 trials that take approximately 1 minute; mean completion time 0.75 seconds, SD 0.57 minutes). In Odd One Out, the participants are presented with 6 symbols and identify which symbol differs from the others as quickly as possible. For example, a trial may have 5 pictures of squares and 1 picture of a rectangle. Participants must identify the rectangle as the correct choice. Each administration contains 9 trials. The task is scored by adding the number of correct responses. This total correct response score is recorded as a measure of working memory. The reaction time in seconds for each trial is also recorded as a secondary domain.

Mobile VLMT

The cognitive domain assessed is recognition memory, and the time to complete is approximately 2.5 minutes in total (approximately 1.5 minutes for short delay, which includes word list presentation, and 1 minute for long delay). Participants are presented with a list of 12 words and given 30 seconds to learn the words. The list is then removed, and participants complete a distractor task (ie, one of the other mobile cognitive

tests). Next, participants are presented with 24 words, which are a mix of the 12 target words and 12 foil words in a recognition memory paradigm. Words are presented one by one, and participants indicate whether the word was on the original list (*yes* or *no*). A total of 14 different lists were presented over the protocol period, and each list was presented only once during each mobile cognitive testing session. The development of the word lists has been described in previous publications [15,45]. The word list recognition trial was always completed within the same survey session (short delay). In addition, 9 times over the 14 days, participants completed a second recognition trial of the list during the subsequent survey as a measure of delayed recognition. This second list recognition task (long delay) always occurred on the same day as the list presentation. The overall score is calculated as a sum of the number of target words correctly identified and the number of foil words correctly rejected, with a total possible score of 24. A modified version of the VLMT in a different sample of persons with serious mental illness was described by Parrish et al [15].

Quick Tap 1

The cognitive domain assessed is processing speed, and the time to complete is approximately 1 minute in total for 12 trials. In Quick Tap 1, participants are first presented with a gray square containing an image (eg, a cartoon dog) that is indicated as the *target*. Participants are asked to tap the target image as quickly as possible when it appears. At the start of a trial, the gray square says, *Wait for the target*. Then, the target image replaces this text in a randomly generated time interval of 1 to 5 seconds, and participants are asked to tap the target as quickly as they can. If the image is not selected, the trial times out 2 seconds after the target is displayed, and the response is marked as incorrect. Similarly, if the gray square is pressed before the target is shown, the response is incorrect, and the next trial begins. Each session contains 12 trials, and each session contains a different target image. This task is scored by averaging the reaction time of the correctly answered trials in seconds.

Quick Tap 2

The cognitive domain assessed is response inhibition, and the time to complete is approximately 1 minute in total for 12 trials. Quick Tap 2 is similar to Quick Tap 1 except that, in some instances, a foil or *trick* (eg, cartoon picture of a cat) image is presented instead of the target image. This foil image, although different, appears similar to the target image (eg, a cartoon boy smiling vs a cartoon boy with his tongue out), and both the target and trick are identified in the instructions before beginning the session. Each session contains a different target and trick combination. Participants respond by tapping the target when it appears as quickly as possible; they are instructed not to tap the foil. Similar to Quick Tap 1, either image is randomly presented within 1 to 5 seconds in each trial, and the next trial begins either immediately after the image is pressed or after 2 seconds of presentation of the image if the image is not selected. The probability of foil presentation for each trial is randomly generated and ranges from 30% to 60%. Quick Tap 2 always immediately follows Quick Tap 1. Each session contains 12 trials. Both Quick Tap 1 and Quick Tap 2 were presented 9 times each over the 14-day protocol period. For each correct response, either correctly tapping the target or not tapping the trick, the total score increases by 1 for a maximum possible score of 12. Reaction time is also recorded.

CopyKat

The cognitive domain assessed is visual working memory, and the time to complete is variable (3 trials that take approximately 2 to 3 minutes; mean completion time 2.7 minutes). Similar to the popular electronic game Simon, the participants are presented with a 2×2 matrix of colored tiles in a fixed position: red, yellow, blue, and green. The tiles briefly light up in a random order, and participants are asked to replicate the pattern by pressing on the colored tiles in the correct order. The number of tiles that light up begins at 1 and increases by 1 with each correct response. For a correct response, the next trial contains the same pattern as the previous trial plus 1 additional highlighted tile. When an incorrect response is made, the same sequence is presented again. Similarly, if no response is made after 20 seconds, the trial is marked as incorrect. There is no upper limit on the maximum number of tiles. The session ends after 3 incorrect responses. The task is scored by summing the number of correct trials. An additional task feature is that each color plays a distinguishable tone when highlighted if the phone volume is turned on.

Statistical Analyses

To compare baseline demographics and clinical characteristics by diagnostic status, independent 2-tailed *t* tests or chi-square tests were used for continuous and categorical variables, respectively. Adherence was examined by calculating the percentage of completed tests for each participant. Average adherence was compared by diagnostic status using independent *t* tests. To understand what we call *fatigue effects* (ie, the likelihood of missing a test), a mixed effect logistic regression was used to examine whether the participants were more likely to miss a mobile cognitive test on later versus earlier study days. An interaction between diagnostic status and study day was included to examine whether fatigue effects differed by diagnosis. The scores on 71% (5/7) of the tests were normally distributed (Multimedia Appendix 1). Odd One Out (total score) had a restricted range of performance, approaching a ceiling effect. Quick Tap 2 had 4 outliers, resulting in a significant skew. We chose not to transform these raw data into analyses.

We created a composite score for the mobile cognitive tests excluding the Odd One Out total score and Quick Tap 2 owing to ceiling effects. To create this composite score, we first log-transformed variables that were positively skewed (CopyKat time, Odd One Out time, Quick Tap 1 average time, and CopyKat total) and squared negatively skewed variables (Matching Pair, Memory Matrix, and VLMT) to normalize them before creating the composite. We then created standardized *Z* scores from these variables, and these standardized *Z* scores were aggregated by participant and combined to create the composite score.

The intraclass correlation coefficients by group are presented in Multimedia Appendix 2. Intraindividual variability was calculated using the mean square of successive differences, which is the sum of the squared differences between 2

consecutive observations divided by 2 times the number of observations minus 1. An independent sample *t* test was used to determine if intraindividual variability differed between participants with BD and healthy volunteers. Performance on each mobile cognitive test was aggregated within each participant across all administrations to examine the average differences in performance by diagnostic group, demographics, and phone type, as well as to examine convergent validity with laboratory-based cognitive tests. Independent *t* tests, ANOVAs with follow-up Tukey HSD pairwise comparisons, and Pearson *r* correlations were used for dichotomous (eg, diagnostic group), >2-level categorical (eg, phone type), and continuous variables (eg, age), respectively.

Linear mixed effect regression models examined the practice effects for each mobile cognitive test to understand whether performance improved as a function of study day. Mixed effect Poisson regression was used for each mobile cognitive test outcome (ie, Odd One Out total score). An interaction with diagnostic status was initially included in each mixed effect model to examine whether practice effects differed by diagnostic status. If the interaction was not significant at $P<.05$, then diagnostic status and its interaction with the study day were removed from the model to simply estimate practice effects in the overall sample. When significant practice effects were identified, spline regression models were conducted to determine whether there was a point (study day) at which the performance stabilized. All statistical analyses were performed using R (version 3.5.0; R Foundation for Statistical Computing). Mixed effect models were conducted using the *lme4* package [46].

Public Significance Statement

There are several limitations to traditional tests of cognitive abilities, including the time, cost, and accessibility of neuropsychological services. This study provides initial research support for 7 newly developed mobile cognitive tests that can be easily self-administered on personal smartphones in real-world environments.

Results

Sample Characteristics

Demographic and clinical characteristics by diagnostic status are shown in Table 3. The groups were comparable in terms of demographics; they did not significantly differ in age, sex, race, ethnicity, or years of education. Participants with BD were more likely than healthy volunteers to be unemployed (23/45, 51% vs 1/21, 5%, respectively) and have a lower income. In addition, compared with healthy volunteers, participants with BD had significantly greater depressive symptomology and lower NIH-TB-CB Fluid Cognition, executive functions, and episodic memory on laboratory-based neuropsychological tests. However, cognitive performance in the BD group was still within normal limits, and the average mania severity scores were within the mild range. A total of 41% (27/66) of the participants used personal iPhones, 32% (21/66) used personal Android phones, and 27% (18/66) used study iPhones.

Table 3. Demographics and clinical characteristics by bipolar disorder status (N=66).

Characteristics	Bipolar disorder group (n=45)		Healthy volunteer group (n=21)		Cohen *d*	Test statistics[a]	*P* value
	Value, mean (SD)	Range or %	Value, mean (SD)	Range or %			
Demographics							
Age (years)	43 (12)	19-61	42 (14)	18-65	0.09	t_{64}=−0.36	.72
Sex (women)	30 (N/A[b])	66.7	15 (N/A)	71.4	N/A	X^2_1=0.2	.70
Race, n (%)					N/A	X^2_3=6.2	.11
White	26 (N/A)	57.8	8 (N/A)	38.1			
Black or African American	4 (N/A)	8.9	2 (N/A)	9.5			
Asian	2 (N/A)	4.4	5 (N/A)	23.8			
Other	13 (N/A)	28.9	6 (N/A)	28.6			
Ethnicity (Hispanic or Latino), n (%)	8 (N/A)	N/A	3 (N/A)	14.3	N/A	X^2_1=0.2	.70
Education (years), mean (SD)	14.89 (N/A)	2.54	15.52 (N/A)	2.73	0.25	t_{64}=0.92	.36
Employment status, n (%)					N/A	X^2_2=12.73	.005[c]
Unemployed	22 (N/A)	48.9	1 (N/A)	4.8			
In school	1 (N/A)	2.2	1 (N/A)	4.8			
Part-time employment	9 (N/A)	20	6 (N/A)	28.6			
Full-time employment	13 (N/A)	28.9	13 (N/A)	61.9			
Residential status, n (%)					N/A	X^2_2=0.5	.79
Independent, financially responsible	36 (N/A)	80	17 (N/A)	81			
Independent, not financially responsible	8 (N/A)	17.8	4 (N/A)	19			
Unsupervised residential facility	0 (N/A)	0	0 (N/A)	0			
Supervised residential facility	1 (N/A)	2.2	0 (N/A)	0			
Income (US $), n (%)					N/A	X^2_2=7.7	.02[d]
<19,000	24 (N/A)	53.3	4 (N/A)	19			
20,000 to 74,999	12 (N/A)	26.7	12 (N/A)	57.1			
>75,000	9 (N/A)	20	5 (N/A)	23.8			
Smartphone used for study, n (%)					N/A	X^2_2=6.4	.04[d]
Personal iPhone	14 (N/A)	31.1	13 (N/A)	61.9			
Personal Android	18 (N/A)	40	3 (N/A)	14.3			
Study-loaned phone	13 (N/A)	28.9	5 (N/A)	23.8			
Substance use and mood							
Alcohol, n (%)					N/A	X^2_2=4.2	.13
Abstinent	21 (N/A)	46.7	6 (N/A)	28.6			
Infrequent-moderate	21 (N/A)	46.7	15 (N/A)	71.4			
Heavy or very heavy	3 (N/A)	6.7	0 (N/A)	0			
Cannabis, n (%)					N/A	X^2_2=6.2	.10

Characteristics	Bipolar disorder group (n=45)		Healthy volunteer group (n=21)		Cohen *d*	Test statistics[a]	*P* value
	Value, mean (SD)	Range or %	Value, mean (SD)	Range or %			
Current abuse	2 (N/A)	4.4	0 (N/A)	0			
Current dependence	2 (N/A)	4.4	0 (N/A)	0			
Former use disorder	7 (N/A)	15.6	0 (N/A)	0			
Beck Depression Inventory-II, mean (SD)	14.69 (9.80)	N/A	2.71 (3.33)	N/A	1.64	$t_{64}=-7.3$	<.001[c]
Baseline YMRS[e], mean (SD)	6.13 (5.32)	N/A	N/A	N/A	N/A	N/A	N/A
Laboratory-based neuropsychological scores[f], mean (SD)							
NIH-TB[g] Total Cognition Score	101.53 (17.13)	N/A	107.38 (11.32)	N/A	0.40	$t_{64}=1.42$	.16
NIH-TB Crystallized Intelligence Score	104.58 (14.69)	N/A	104.95 (14.27)	N/A	0.03	$t_{64}=0.10$	.92
NIH-TB Fluid Cognition Score	98.07 (19.11)	N/A	107.48 (11.29)	N/A	0.60	$t_{64}=2.09$	.04[d]
NIH-TB Flanker Inhibitory Control and Attention Test	91.89 (15.17)	N/A	93.38 (10.32)	N/A	0.11	$t_{64}=0.47$	.64
NIH-TB Dimensional Change Card Sort Test	101.31 (20.25)	N/A	110.38 (11.86)	N/A	0.55	$t_{64}=2.28$	.03[d]
NIH-TB Pattern Comparison Processing Speed Test	103.53 (20.13)	N/A	109.57 (17.06)	N/A	0.32	$t_{64}=1.19$	.24
NIH-TB List Sorting Working Memory Test	97.84 (15.43)	N/A	105.10 (12.87)	N/A	0.51	$t_{64}=1.87$	.07
NIH-TB Oral Reading Recognition Test	101.84 (13.18)	N/A	102.48 (13.60)	N/A	0.05	$t_{64}=0.18$	.86
NIH-TB Picture Sequence Memory Test	99.42 (17.33)	N/A	107.43 (12.54)	N/A	0.53	$t_{64}=2.13$	.04[d]
D-KEFS[h] Color-Word Interference Test	10.82 (3.35)	N/A	12.48 (1.44)	N/A	0.64	$t_{64}=2.80$	.007[c]

[a]*t* tests for continuous variables and chi-square tests for dichotomous variables.

[b]N/A: not applicable.

[c]*P*<.01.

[d]*P*<.05.

[e]YMRS: Young Mania Rating Scale.

[f]Demographically adjusted standard scores from the National Institutes of Health Toolbox Cognition Battery unless otherwise noted.

[g]NIH-TB: National Institutes of Health Toolbox.

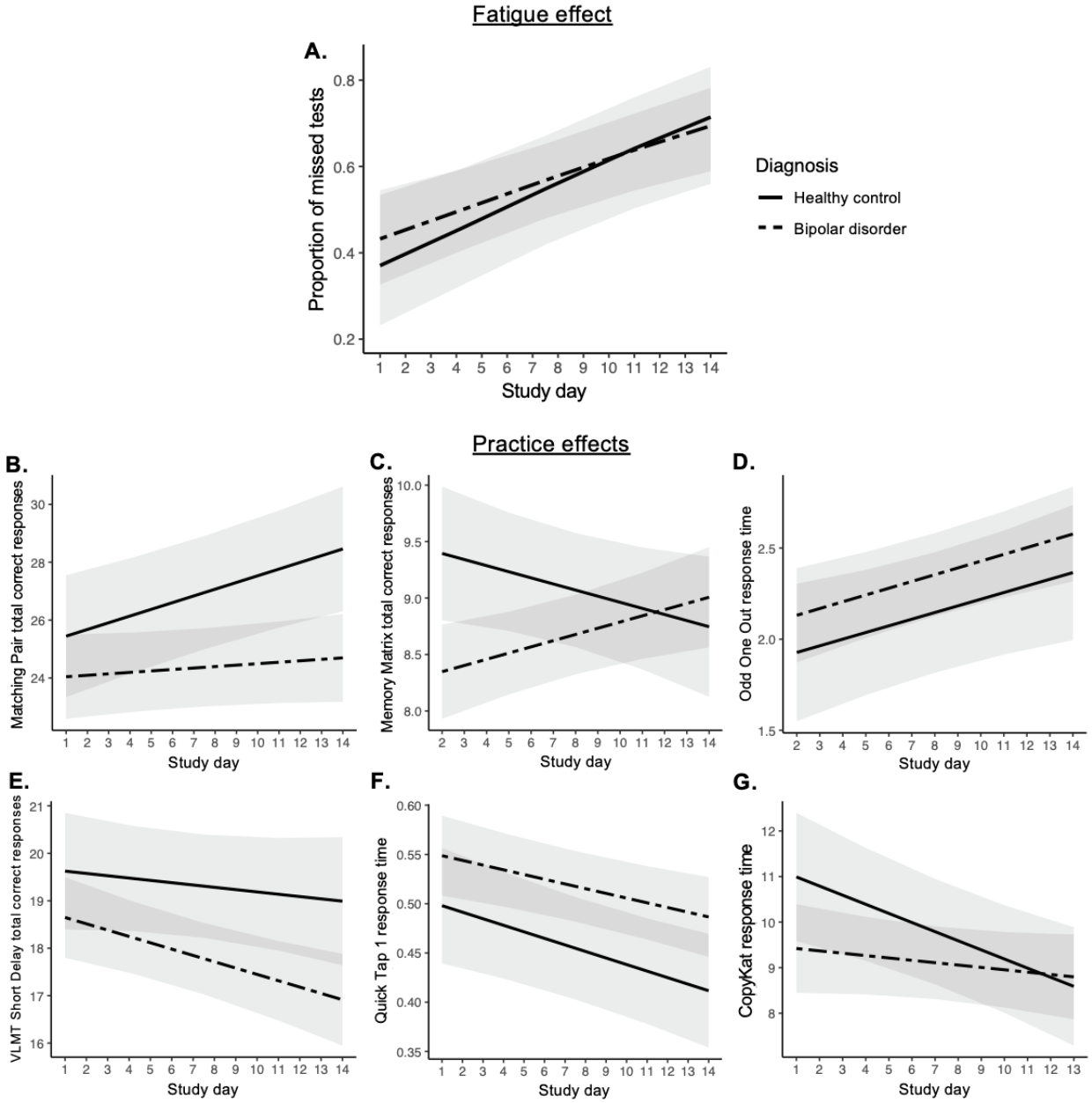
[h]D-KEFS: Delis-Kaplan Executive Function System.

Overall Adherence

Overall, the mean adherence to the EMCT protocol was 69.7% (SD 20.5%, range 28.6%-97.6%), resulting in 3965 valid and complete mobile cognitive tests among all 66 participants. Adherence did not differ by diagnostic status ($t_{64}=0.97$; $P=.33$).

In the overall sample, participants were more likely to miss a test as the number of study days increased (logit=0.093; odds ratio 1.100 [per 1-day increase], 95% CI 1.060-1.138; $P<.001$). This fatigue effect did not differ by diagnostic status ($P=.45$; Figure 1A).

Figure 1. Fatigue effect and practice effects in mobile cognitive test completion over the 14-day study period. VLMT: Variable Difficulty List Memory Test.



Mobile Cognitive Test Performance—Group, Demographic, and Phone Type Differences

Participants with BD were slower than the healthy volunteers on mean Quick Tap 1 performance; mean performance did not differ on the other mobile cognitive tests, although medium effect sizes were found for Matching Pair, Memory Matrix, and the VLMT (Table 4). In each case, the healthy volunteer group outperformed the BD group. Associations between mobile cognitive test performance and age, sex, education, and phone type are presented in Multimedia Appendix 3. In the full sample, age effects were found for Matching Pair, Memory Matrix, Odd One Out (response time), and Quick Tap 1 and 2; all findings were in the expected direction, with worse performance associated with older age. No sex effects were observed in this sample. Higher education was associated with better performance on all tests except Odd One Out (total score) and Quick Tap 2. Participants with personal iPhones had significantly higher scores on all tests (except Quick Tap 2) compared with participants with personal Androids or laboratory-given iPhones. Regarding ethnicity, there were no significant differences in mobile cognitive test performance between Hispanic or Latino and non-Hispanic or Latino participants with BD (P>.05 in all cases). Among healthy volunteers, non-Hispanic or Latino participants had a greater Odd One Out score (mean 8.47, SD 0.37) than Hispanic or Latino participants (mean 7.7, SD 0.90; t_{19}=2.56; P=.02). Non-Hispanic or Latino healthy volunteers had a faster CopyKat reaction time (mean 9.33, SD 1.84) than Hispanic or Latino healthy volunteers (mean 12.09, SD 2.26; t_{19}=−2.34; P=.03). There were no other significant differences in mobile cognitive test performance between Hispanic or Latino and non-Hispanic participants.

Table 4. Mobile cognitive test performance by bipolar disorder status (N=66).

Cognitive domain and individual tests	Bipolar disorder group (n=45), mean (SD; range)	Healthy volunteer group (n=21), mean (SD; range)	Cohen *d*	T-score	*P* value
Processing speed					
Matching Pair (total score)	24.21 (5.23; 16.00-37.90)	26.72 (4.09; 18.67-34.60)	0.53	1.94	.06
Odd One Out (response time)	2.27 (0.81; 1.15-5.59)	2.08 (0.50; 1.36-3.20)	0.28	−1.01	.32
Quick Tap 1 (response time)	0.51 (0.14; 0.34-0.91)	0.45 (0.07; 0.35-0.65)	0.54	−1.89[a]	.02[b]
Working memory					
Memory Matrix (total score)	8.59 (1.35; 5.50-11.00)	9.11 (0.82; 7.38-10.88)	0.47	1.92[a]	.06
Odd One Out (total score)	8.2 (0.58; 5.67-8.89)	8.36 (0.51; 7.20-9.00)	0.29	1.08	.28
CopyKat (total score)	9.08 (3.45; 2.60-19.33)	10.27 (2.52; 5.00-15.63)	0.39	1.41	.16
Recognition memory—VLMT[c] (total score)	16.97 (2.35; 12.96-21.87)	18.2 (2.32; 12.79-23.37)	0.53	1.98	.05
Response inhibition—Quick Tap 2 (total score)	10.74 (1.14; 7.00-12.00)	11.03 (1.19; 6.33-11.89)	0.25	0.96	.34

[a]Levene test for equality of variances violated; equal variances were not assumed.

[b]*P*<.05.

[c]VLMT: Variable Difficulty List Memory Test.

Mobile Cognitive Test Adherence and Performance

Overview

Correlations between average mobile cognitive test scores and laboratory-based cognitive performance are presented in Table 5 (BD participants) and Table 6 (healthy volunteers). The mobile cognitive tests were designed to measure fluid cognition (vs crystallized intelligence). In most cases, the average mobile cognitive test performance and composite score were moderately to strongly correlated with the validated NIH-TB-CB fluid measures (age-corrected Fluid Cognition Composite Score and individual test scores) and the D-KEFS Color-Word Interference Test (*P*<.001 in each case). Adherence, practice effects, convergent validity with the NIH-TB Fluid Cognition Composite Score and individual tests, convergent validity with the D-KEFS Color-Word Interference Test, and intraindividual variability are provided below.

Table 5. Correlations between mobile cognitive tests and in-laboratory neuropsychological performance in the bipolar disorder sample (N=45).

	NIH[a] Toolbox Cognition Battery[b]						D-KEFS[c] Color-Word Interference Test
	Fluid Cognition Composite Score	Flanker Inhibitory Control and Attention Test	Dimensional Change Card Sort Test	Pattern Comparison Processing Speed Test	List Sorting Working Memory Test	Picture Sequence Memory Test	
Processing speed							
Matching Pair	0.637[d]	0.302[e]	0.488[d]	0.497[d]	0.518[d]	0.469[d]	0.592[d]
Odd One Out (response time)	−0.537[d]	−0.385[e]	−0.439[d]	−0.419[d]	−0.327[e]	−0.354[e]	−0.220
Quick Tap 1	−0.575[d]	−0.448[e]	−0.522[d]	−0.471[d]	−0.327[e]	−0.290	−0.364[d]
Working memory							
Memory Matrix	0.582[d]	0.217	0.509[d]	0.430[d]	0.495[d]	0.421[d]	0.431[d]
Odd One Out (total score)	0.154	−0.096	0.117	0.043	0.329[e]	0.168	0.285
CopyKat	0.537[d]	0.234	0.460[d]	0.408[d]	0.383[d]	0.419[d]	0.564[d]
Recognition memory							
VLMT[f]	0.213	−0.072	0.228	0.047	0.336[e]	0.224	0.535[d]
Response inhibition							
Quick Tap 2	−0.083	−0.178	−0.069	0.075	0.115	−0.251	0.335[e]
Composite score	0.494[d]	0.072	0.379[e]	0.352[e]	0.526[d]	0.427[d]	0.664[d]

[a]NIH: National Institutes of Health.

[b]Age-corrected standard scores; lower scores indicate a slower (worse) performance, and higher scores indicate a better performance.

[c]D-KEFS: Delis-Kaplan Executive Function System.

[d]*P*<.01.

[e]*P*<.05.

[f]VLMT: Variable Difficulty List Memory Test.

Table 6. Correlations between mobile cognitive tests and in-laboratory neuropsychological performance in the healthy control sample (N=21).

	NIH[a] Toolbox Cognition Battery[b]						D-KEFS[c] Color-Word Interference Test
	Fluid Cognition Composite Score	Flanker Inhibitory Control and Attention Test	Dimensional Change Card Sort Test	Pattern Comparison Processing Speed Test	List Sorting Working Memory Test	Picture Sequence Memory Test	
Processing speed							
Matching Pair	0.483[d]	−0.08	−0.357	0.540[d]	0.641[e]	0.451[d]	−0.101
Odd One Out (response time in seconds)	−0.317	0.076	0.286	−0.301	−0.406	−0.439[d]	0.355
Quick Tap 1	−0.459[e]	−0.011	−0.048	−0.454[e]	−0.100	−0.600[e]	−0.081
Working memory							
Memory Matrix	0.540[d]	0.077	−0.041	0.595[e]	0.38	0.394	0.077
Odd One Out (total score)	0.496[d]	0.199	0.151	0.373	0.266	0.389	0.236
CopyKat	0.376	−0.088	−0.025	0.411	0.349	0.314	0.178
R ecognition memory							
VLMT[f]	0.428	−0.059	−0.024	0.277	0.535[d]	0.414	−0.345
Response inhibition							
Quick Tap 2	0.001	−0.093	−0.15	0.205	0.067	−0.11	0.223
Composite score	0.478[d]	−0.021	0.175	0.511[d]	0.608[e]	0.292	0.015

[a]NIH: National Institutes of Health.

[b]Age-corrected standard scores; lower scores indicate slower (worse) performance, and higher scores indicate a better performance.

[c]D-KEFS: Delis-Kaplan Executive Function System.

[d]$P<.05$.

[e]$P<.01$.

[f]VLMT: Variable Difficulty List Memory Test.

Matching Pair

The mean adherence to Matching Pair was 75.3% (SD 23.7%, range 11.1%-100%), and adherence did not differ by diagnostic status (t_{64}=0.96; P=.34). There was a significant difference in practice effect by diagnostic status (estimate=−0.18; SE 0.09; P=.04) such that scores increased within persons across days among healthy volunteers (estimate=0.23; SE 0.08; P=.003) but not among persons with BD (estimate=0.05; SE 0.05; P=.32; Figure 1B). Spline regressions did not identify any point at which Matching Pair performance stabilized among the healthy volunteers. In both groups, Matching Pair was moderately to strongly associated with the NIH-TB Fluid Cognition Composite Score (BD participants: r=0.64 and P<.001; healthy volunteer participants: r=0.48 and P=.03). Intraindividual variability did not differ by diagnostic status (t_{64}=0.78; P=.44).

Memory Matrix

The mean adherence in the overall sample to Memory Matrix was 75.3% (SD 21.3%, range 22.2%-100%). Participants with BD had lower adherence than healthy volunteers (292/405, 72.1% vs 155/189, 82%), but the difference was not statistically significant (t_{64}=1.79; P=.08). There was a significant difference in practice effect by diagnostic status (estimate=0.11; SE 0.04; P=.002) such that (in contrast to Matching Pair) scores increased within persons among those with BD (estimate=0.05; SE 0.02; P=.005) but not among healthy volunteers (estimate=−0.05; SE 0.03; P=.12; Figure 1C). Spline regression did not identify any point at which Memory Matrix performance stabilized in the BD group. Memory Matrix was strongly related to the NIH-TB Fluid Cognition Composite Score in both groups (BD participants: r=0.58 and P<.001; healthy volunteer participants: r=0.54 and P=.01). Intraindividual variability did not differ by diagnostic status (t_{64}=0.54; P=.59).

Odd One Out

The mean adherence in the overall sample to the Odd One Out mobile cognitive test was 69.9% (SD 21.8%, range 11.1%-100%), and adherence did not differ by diagnostic status (t_{64}=0.67; P=.51). For the Odd One Out total correct variable, there was no practice effect in the overall sample (P=.15), nor was there a difference by diagnostic status (P=.80). For the Odd One Out response time variable, participants' average response times increased across days in the overall sample, indicating

worsening performance over time (estimate=0.037; SE 0.01; P<.001). This effect did not differ by diagnostic status (P=.97). Item difficulty did not vary across Odd One Out trials. There was no difference in this effect by diagnostic status (P=.97; Figure 1D). Spline regressions identified that the Odd One Out average response time increased significantly until day 4 (the second remote administration; days 1-4 regression estimate=0.42; SE 0.05; P<.001), after which response times stabilized (days 4-14 regression estimate=−0.01; SE 0.01; P=.30). The Odd One Out total correct score was unrelated to the NIH-TB Fluid Cognition Composite Score in the BD group (r=0.15; P=.36), but these variables were strongly related in the healthy volunteer group (r=0.50; P=.02). Conversely, the Odd One Out response time was strongly associated with the NIH-TB Fluid Cognition Composite Score in the BD group (r=−0.54; P<.001), but these variables were unrelated in the healthy volunteer group (r=−0.32; P=.13). Intraindividual variability of Odd One Out total correct score (t_{64}=0.57; P=.57) and response time (t_{64}=−0.90; P=.37) did not differ by diagnostic status.

Mobile VLMT

The mean adherence to the VLMT short delay in the overall sample was 66.5% (SD 18.9%, range 18.2%-100%). Adherence did not differ by diagnostic status (t_{64}=0.74; P=.46). For the VLMT long delay, the mean adherence was 63.5% (SD 27.6%, range 0%-100%), which again did not differ by diagnostic status (t_{64}=−0.38; P=.70). In the overall sample, although there was no practice effect for the short delay, participants appeared to recall fewer total correct words at the short delay over time (estimate=−0.10; SE 0.03; P=.002). This effect did not differ by diagnostic status (P=.21; Figure 1E). Spline regressions identified that short-delay scores decreased significantly until day 7 (seventh remote administration; days 1-7 regression estimate=−0.13; SE 0.07; P=.05), after which the scores stabilized (days 7-14 regression estimate=−0.07; SE 0.07; P=.31). For the VLMT long delay, there was no significant practice effect for the VLMT total correct score in the overall sample (estimate=−0.03; SE 0.05; P=.61), nor did this effect differ by diagnostic status (P=.24). The VLMT was not significantly related to the NIH-TB Fluid Cognition Composite Score in either group (BD participants: r=0.21 and P=.16; healthy volunteer participants: r=0.43 and P=.05). This shows good discriminant validity given that the NIH-TB-CB does not include a verbal recognition test. Intraindividual variability did not differ by diagnostic status (t_{64}=0.79; P=.44).

We also examined within-person forgetting on the VLMT. From short to long delay, participants lost an average of 3.8 words (SD 2.4, range −5 to 12; negative values reflect words gained from short to long delay). The average number of words lost from short to long delay was not related to diagnostic status (healthy volunteer mean 4.6 words, BD mean 3.4 words; t_{62}=1.8; P=.07) or demographic characteristics of the participants, including age (b=−0.00; SE 0.025; P=.90), sex (b=−0.01; SE 0.65; P=.88), years of education (b=−0.02; SE 0.12; P=.89), ethnicity (Hispanic vs non-Hispanic; b=0.79; SE 0.80; P=.33), or White (vs other) race (b=−0.59; SE 0.60; P=.33).

Quick Tap 1

The mean adherence to Quick Tap 1 was 72.4% in the overall sample (SD 20.9%, range 22.2%-100%), and adherence did not differ by diagnostic status (t_{64}=1.01; P=.32). In the overall sample, there was a slight but significant practice effect for Quick Tap 1 response time. Response times decreased within persons across days (estimate=−0.01; SE 0.00; P<.001). This effect did not differ by diagnostic status (P=.46; Figure 1F). Spline regressions identified that response times were stable from day 1 to day 3 (days 1-3 regression estimate=0.01; SE 0.01; P=.27) and then significantly decreased after day 3 (days 3-14 regression estimate=−0.01; SE 0.00; P<.001). Quick Tap 1 was strongly related to the NIH-TB Fluid Cognition Composite Score in both groups (BD participants: r=−0.58 and P<.001; healthy volunteer participants: r=−0.46 and P=.04). Intraindividual variability did not differ by diagnostic status (t_{64}=−0.75; P=.46).

Quick Tap 2

Participants were 72.2% adherent to Quick Tap 2 on average (SD 21.1%, range 22.2%-100%); this did not differ by diagnostic status (t_{64}=1.04; P=.30). In the overall sample, there was no practice effect for the Quick Tap 2 total correct response score (estimate=−0.00; SE 0.01; P=.76), and there was no difference in practice effect by diagnostic status (P=.63). Quick Tap 2 scores were unrelated to the NIH-TB Fluid Cognition Composite Score in both groups (BD participants: r=−0.08 and P=.59; healthy volunteer participants: r=0.001 and P=.99). Intraindividual variability did not differ by diagnostic status (t_{64}=0.24; P=.81).

CopyKat

In the overall sample, participants were 77.9% adherent to CopyKat on average (SD 21.6%, range 22.2%-100%). Adherence did not differ by diagnostic status (t_{64}=1.19; P=.24). In the overall sample, there was no practice effect for the CopyKat total correct response score (estimate=−0.07; SE 0.04; P=.13), and there was no difference by diagnostic status (P=.13). However, for CopyKat average reaction time, there was a significant difference in practice effect by diagnostic status (estimate=0.15; SE 0.07; P=.05) such that reaction time significantly decreased (ie, improved) among healthy volunteers over time (estimate=−0.20; SE 0.06; P=.001) but not among participants with BD (estimate=−0.05; SE 0.04; P=.24; Figure 1G). Spline regressions identified that, among healthy volunteers, average reaction times decreased significantly until day 9 performance (sixth administration; days 1-9 regression estimate=−0.20; SE 0.10; P=.05), after which reaction times stabilized (estimate=−0.20; SE 0.18; P=.27). The CopyKat total correct response score was strongly related to the NIH-TB Fluid Cognition Composite Score in BD participants (r=0.54; P<.001); these variables were unrelated in the healthy volunteer group (r=0.38; P=.09). Intraindividual variability for the CopyKat total correct response score did not differ by diagnostic status (t_{64}=1.83; P=.08).

Discussion

Principal Findings

The findings of this study support the acceptability and preliminary psychometric properties of 7 brief, repeatable, newly developed mobile cognitive tests assessing processing speed, reaction time, visual working memory, recognition memory, and response inhibition in people with BD as well as in a small sample of healthy volunteers. The test stimuli can be used on personal smartphones and, thus, were designed with a simple user interface to accommodate a diversity of human-level factors (eg, varying cognitive abilities and range of sociodemographic groups) as well as a diversity of device factors (eg, different operating systems and screen sizes). Given that these tests are intended to be self-administered, a combination of clear verbal and visual instructions was incorporated. The main findings of this study, which were broadly consistent with our hypotheses, include (1) adequate adherence to the study protocol, with participants completing an average of 70% of the 42 EMCT administrations over a 14-day period, and no group differences in adherence; (2) differences in sociodemographic factors and mobile cognitive test performance based on ownership of an iPhone versus an Android device [47]; (3) moderate to strong correlations between the mobile cognitive tests and in-laboratory neuropsychological performance in the whole sample; (4) a fatigue effect such that participants were more likely to miss tests as the number of study days increased (with no differences between the BD and healthy volunteer groups); and (5) small practice effects, primarily among tests assessing response time. This study adds to the growing literature supporting the convergent and discriminant validity of mobile cognitive testing by demonstrating greater degrees of shared variance between mobile and laboratory-based tests of similar constructs than between mobile and laboratory-based tests of disparate constructs (Tables 4 and 5) [13,14,16,45,48].

The only test in which we found a statistically significant difference between the groups was Quick Tap 1. In both groups, low intraindividual variability was observed in Quick Tap 1; this test may be a good candidate for studies or clinical trials looking for a simple processing speed test that is not overly variable and may be most sensitive to cognitive dysfunction detection. Furthermore, the mobile cognitive tests that showed the most promise for testing real-world cognitive performance among people with BD (based on the examination of effect sizes) were tests tapping the domains of processing speed (Matching Pair and Quick Tap 1), working memory (Memory Matrix), and recognition memory (VLMT). These cognitive domains are associated with disability and mood state effects in persons with BD and, thus, may be useful tools for examining cognition change over time or response to treatment.

Of note, the correlations between our mobile cognitive tests and in-laboratory neuropsychological performance were less consistent in healthy controls than in participants with BD. This is likely related to the smaller sample size of the healthy control group and the more restricted range of scores on both mobile cognitive tests and in-laboratory neuropsychological tests among the healthy controls as limited variability restricts our ability to detect robust relationships between these measures. Future work should continue to validate these mobile tests in larger samples of healthy adults. Another finding we want to make note of is the unexpected results with the Odd One Out response times, in that participants did worse on this task (slower performance) as the number of study days increased. This could be due to fatigue or lower motivation to complete the task quickly after repeated administrations. Anecdotally, participants also reported to the study staff that they found this task "boring," which may have affected test performance with repeated administrations. Nevertheless, adherence to this task was still high at an average of 69.9% (SD 21.8%), indicating acceptability among the participants. We also saw some evidence of ceiling effects for this task as well as Quick Tap 2, and further evaluation and iterations of these tasks are likely warranted to improve psychometrics and increase task engagement. Further work would also benefit from examining whether these tasks could be useful in other ways, such as measuring effort when completing fully remote ambulatory assessments.

This study adds to the limited literature on foundational psychometric research of an EMCT platform. Many of the mobile cognitive tests available for download do not have accompanying psychometric data to guide their use [49]. Previous studies support the feasibility and acceptability of mobile cognitive testing, with adherence rates ranging from approximately 79% to 90% in both clinical and nonclinical samples [14-16,45,50,51]. Most [16,48,52] but not all [13] previous investigations report small practice effects on selected tests, with no differences across clinical and nonclinical groups. A study reported a small impact of fatigue, with reductions in adherence rates across the 14-day procedure [16], whereas 2 other investigations reported no such fatigue effects [48,52].

Limitations and Considerations for Future Research

This study is not without limitations. First, the sample size was relatively small, and the participants were of a specific psychiatric population, with only 32% (21/66) of healthy volunteer participants included. Our examination of intraindividual variability by diagnostic status was underpowered to detect meaningful differences; however, we included these data as they still provide valuable insights into performance in this sample. As previously mentioned, a recently published study provided validity data for the VLMT (6-, 12-, and 18-length word lists) among a large sample of persons with serious mental illness [15]. Further work is ongoing to continue the validation of these tasks and others in different populations, including larger groups of healthy volunteers and persons with mild cognitive impairment. We recommend that this ongoing (and future) work consider traditional sociodemographic factors that are known to affect cognitive test performance (age, sex, race or ethnicity, and education), the amount of variability in test scores that is attributable to device type, and digital literacy. Future work would also benefit from examining whether creating a composite variability index that includes performance on >1 test would yield variability data that are more sensitive than variability on each individual test.

Second, we were unable to examine the convergent validity of our tests with traditional laboratory-based cognitive domain

scores, and the NIH-TB-CB does not include >1 test per domain to generate domain score data. However, we were able to demonstrate the overall construct validity of our mobile cognitive tests and our composite score (calculated using *Z* scores) with the NIH-TB Fluid Composite Score and present relationships with individual NIH-TB tests, providing proof of concept for future studies. Our sample size was insufficient to create demographically corrected T-scores with our mobile cognitive tests, which is the best-practice metric for creating composites [53]; subsequent studies are needed to generate large-scale normative data.

Third, although we examined smartphone type, we did not collect data on service providers (eg, T-Mobile and AT&T), all of which have different data speeds and connectivity, and we did not examine differences by screen size (eg, iPhone 8 vs iPhone 8 Plus), which may affect response times. Similarly, we did not collect smartphone metadata, which could create variance in test performance, such as accelerometry, touch sensitivity and latency (which, although improving in consistency, can differ by up to 100 milliseconds between different devices [54,55]), and frame rate for dynamic visual displays. Ongoing digital phenotyping studies by our group and several other research groups include the collection of both active (eg, surveys and mobile cognitive tests) and passive (eg, accelerometry, geolocation, keystroke data, and ambient noise levels) digital data, which can provide a more comprehensive picture of cognition in context as well as allow for the examination of the aforementioned limitations. Refer to Germine et al [34] for an excellent review of the challenges faced when digitizing neuropsychological testing as well as a road map of specific recommendations. In line with these recommendations, the approaches we have taken in this study include taking a user-centered design approach to developing tests and developing a flexible platform that can be modified to accommodate changes in technology and adapted or customized to individual investigator requirements.

A final limitation that applies to all remote mobile cognitive testing is that it is difficult to identify suspected cheating, such as whether the participant or someone else took the tests. One way to address this is to examine score distributions and flag outliers as potential instances of noncompliance. Other options would be to include the collection of biodata such as fingerprints or face IDs. Relatedly, it is difficult to assess effort in a mobile cognitive testing platform. Previous work has found a small effect of self-reported distractions and interruptions on mobile cognitive testing performance but also that convergent validity with laboratory-based tests was minimally affected by these factors [15]. Another potential indicator is a lack of a practice effect on a test in which a practice effect is expected.

Conclusions

In conclusion, the 7 mobile cognitive tests we have presented in this study may serve as useful tools for brief, frequent ambulatory cognitive testing in a person's everyday environment. Our data show that people with a well-characterized psychiatric disorder can and will complete self-administered mobile cognitive tests with good adherence. The tests are automatically scored, can be integrated with EMA surveys, and are available for other investigators to use; thereby, they are poised for further psychometric (including norming) studies and scalability. As the field of neuropsychology continues to migrate toward precision medicine, there are several advantages to the availability of psychometrically strong mobile cognitive tests, including complementing traditional neuropsychological assessments by gathering data outside of the controlled clinic environment, examining intraindividual variability and establishing more reliable estimates of cognitive performance over time, improving sensitivity to detect change and reducing the number needed to treat in clinical trials, and ultimately having the capability to detect brain dysfunction and risk of cognitive decline earlier than is possible with traditional assessment methods.

Acknowledgments

This research was supported by grant R21 MH116104 from the National Institute of Mental Health. JB's time was supported by a Veterans Administration Clinical Science Research and Development grant CX001600. LTE was supported by the Desert-Pacific Mental Illness Research Education and Clinical Center and grant R01MH103318.

Conflicts of Interest

RCM is a cofounder of KeyWise AI Inc, and a consultant for NeuroUX. The terms of this arrangement have been reviewed and approved by the University of California San Diego in accordance with its conflict-of-interest policies. DL is the owner and operator of Playpower Labs Inc, the company that developed the NeuroUX software used in the study.

Multimedia Appendix 1
Skewness and kurtosis of aggregated mobile cognitive testing variables.
[DOCX File , 16 KB - jmir_v24i7e36665_app1.docx]

Multimedia Appendix 2
Intraclass correlation coefficient between participant groups.
[DOCX File , 15 KB - jmir_v24i7e36665_app2.docx]

Multimedia Appendix 3
Correlations among mobile cognitive tests, demographics, and phone type.
[DOCX File , 20 KB - jmir_v24i7e36665_app3.docx]

References

1. Internet World stats: usage and population statistics. Internet World Stats. URL: https://www.internetworldstats.com/stats.htm2021 [accessed 2022-10-01]
2. Mobile fact sheet. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/#who-owns-cellphones-and-smartphones: [accessed 2022-10-01]
3. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008 Apr 01;4(1):1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
4. Bell IH, Lim MH, Rossell SL, Thomas N. Ecological momentary assessment and intervention in the treatment of psychotic disorders: a systematic review. Psychiatr Serv 2017 Nov 01;68(11):1172-1181. [doi: 10.1176/appi.ps.201600523] [Medline: 28669284]
5. Burke LE, Shiffman S, Music E, Styn MA, Kriska A, Smailagic A, et al. Ecological momentary assessment in behavioral research: addressing technological and human participant challenges. J Med Internet Res 2017 Mar 15;19(3):e77 [FREE Full text] [doi: 10.2196/jmir.7138] [Medline: 28298264]
6. May M, Junghaenel DU, Ono M, Stone AA, Schneider S. Ecological momentary assessment methodology in chronic pain research: a systematic review. J Pain 2018 Jul;19(7):699-716 [FREE Full text] [doi: 10.1016/j.jpain.2018.01.006] [Medline: 29371113]
7. Serre F, Fatseas M, Swendsen J, Auriacombe M. Ecological momentary assessment in the investigation of craving and substance use in daily life: a systematic review. Drug Alcohol Depend 2015 Mar 01;148:1-20. [doi: 10.1016/j.drugalcdep.2014.12.024] [Medline: 25637078]
8. Santangelo P, Bohus M, Ebner-Priemer UW. Ecological momentary assessment in borderline personality disorder: a review of recent findings and methodological challenges. J Pers Disord 2014 Aug;28(4):555-576. [doi: 10.1521/pedi_2012_26_067] [Medline: 22984853]
9. Walz LC, Nauta MH, Aan Het Rot M. Experience sampling and ecological momentary assessment for studying the daily lives of patients with anxiety disorders: a systematic review. J Anxiety Disord 2014 Dec;28(8):925-937. [doi: 10.1016/j.janxdis.2014.09.022] [Medline: 25445083]
10. Moore RC, Depp CA, Wetherell JL, Lenze EJ. Ecological momentary assessment versus standard assessment instruments for measuring mindfulness, depressed mood, and anxiety among older adults. J Psychiatr Res 2016 Apr;75:116-123 [FREE Full text] [doi: 10.1016/j.jpsychires.2016.01.011] [Medline: 26851494]
11. Weizenbaum E, Torous J, Fulford D. Cognition in context: understanding the everyday predictors of cognitive performance in a new era of measurement. JMIR Mhealth Uhealth 2020 Jul 23;8(7):e14328 [FREE Full text] [doi: 10.2196/14328] [Medline: 32706680]
12. Allard M, Husky M, Catheline G, Pelletier A, Dilharreguy B, Amieva H, et al. Mobile technologies in the early detection of cognitive decline. PLoS One 2014 Dec 23;9(12):e112197 [FREE Full text] [doi: 10.1371/journal.pone.0112197] [Medline: 25536290]
13. Sliwinski MJ, Mogle JA, Hyun J, Munoz E, Smyth JM, Lipton RB. Reliability and validity of ambulatory cognitive assessments. Assessment 2018 Jan 15;25(1):14-30 [FREE Full text] [doi: 10.1177/1073191116643164] [Medline: 27084835]
14. Moore RC, Swendsen J, Depp CA. Applications for self-administered mobile cognitive assessments in clinical research: a systematic review. Int J Methods Psychiatr Res 2017 Dec 31;26(4):e1562 [FREE Full text] [doi: 10.1002/mpr.1562] [Medline: 28370881]
15. Parrish E, Kamarsu S, Harvey P, Pinkham A, Depp C, Moore R. Remote ecological momentary testing of learning and memory in adults with serious mental illness. Schizophr Bull 2021 Apr 29;47(3):740-750 [FREE Full text] [doi: 10.1093/schbul/sbaa172] [Medline: 33219382]
16. Moore R, Campbell L, Delgadillo J, Paolillo E, Sundermann E, Holden J, et al. Smartphone-based measurement of executive function in older adults with and without HIV. Arch Clin Neuropsychol 2020 May 22;35(4):347-357 [FREE Full text] [doi: 10.1093/arclin/acz084] [Medline: 31942632]
17. Koo B, Vizer L. Mobile technology for cognitive assessment of older adults: a scoping review. Innov Aging 2019 Jan;3(1):igy038 [FREE Full text] [doi: 10.1093/geroni/igy038] [Medline: 30619948]
18. Swendsen J, Schweitzer P, Moore R. Mobile cognitive testing using experience sampling methodology. In: Experience Sampling in Mental Health Research, 1st edition. New York, NY: Routledge; 2019.
19. American Psychiatric Association. Bipolar and related disorders. In: Diagnostic And Statistical Manual Of Mental Disorders, Fifth Edition. Virginia, United States: American Psychiatric Association; 2013.
20. Burdick KE, Ketter TA, Goldberg JF, Calabrese JR. Assessing cognitive function in bipolar disorder: challenges and recommendations for clinical trial design. J Clin Psychiatry 2015 Mar 25;76(03):e342-e350. [doi: 10.4088/jcp.14cs09399]

21. Depp C, Mausbach B, Harmell A, Savla G, Bowie C, Harvey P, et al. Meta-analysis of the association between cognitive abilities and everyday functioning in bipolar disorder. Bipolar Disord 2012 May;14(3):217-226 [FREE Full text] [doi: 10.1111/j.1399-5618.2012.01011.x] [Medline: 22548895]
22. Depp CA, Moore RC, Dev SI, Mausbach BT, Eyler LT, Granholm EL. The temporal course and clinical correlates of subjective impulsivity in bipolar disorder as revealed through ecological momentary assessment. J Affect Disord 2016 Mar 15;193:145-150 [FREE Full text] [doi: 10.1016/j.jad.2015.12.016] [Medline: 26773907]
23. Depp CA, Ceglowski J, Wang VC, Yaghouti F, Mausbach BT, Thompson WK, et al. Augmenting psychoeducation with a mobile intervention for bipolar disorder: a randomized controlled trial. J Affect Disord 2015 Mar 15;174:23-30 [FREE Full text] [doi: 10.1016/j.jad.2014.10.053] [Medline: 25479050]
24. Husky MM, Gindre C, Mazure CM, Brebant C, Nolen-Hoeksema S, Sanacora G, et al. Computerized ambulatory monitoring in mood disorders: feasibility, compliance, and reactivity. Psychiatry Res 2010 Jul 30;178(2):440-442. [doi: 10.1016/j.psychres.2010.04.045] [Medline: 20488558]
25. Schwartz S, Schultz S, Reider A, Saunders EF. Daily mood monitoring of symptoms using smartphones in bipolar disorder: a pilot study assessing the feasibility of ecological momentary assessment. J Affect Disord 2016 Feb;191:88-93 [FREE Full text] [doi: 10.1016/j.jad.2015.11.013] [Medline: 26655117]
26. Wenze SJ, Armey MF, Miller IW. Feasibility and acceptability of a mobile intervention to improve treatment adherence in bipolar disorder: a pilot study. Behav Modif 2014 Jul 08;38(4):497-515 [FREE Full text] [doi: 10.1177/0145445513518421] [Medline: 24402464]
27. Colombo D, Fernández-Álvarez J, Patané A, Semonella M, Kwiatkowska M, García-Palacios A, et al. Current state and future directions of technology-based ecological momentary assessment and intervention for major depressive disorder: a systematic review. J Clin Med 2019 Apr 05;8(4):465 [FREE Full text] [doi: 10.3390/jcm8040465] [Medline: 30959828]
28. Gershon A, Kaufmann CN, Torous J, Depp C, Ketter TA. Electronic Ecological Momentary Assessment (EMA) in youth with bipolar disorder: demographic and clinical predictors of electronic EMA adherence. J Psychiatr Res 2019 Sep;116:14-18. [doi: 10.1016/j.jpsychires.2019.05.026] [Medline: 31176107]
29. Johnson EI, Grondin O, Barrault M, Faytout M, Helbig S, Husky M, et al. Computerized ambulatory monitoring in psychiatry: a multi-site collaborative study of acceptability, compliance, and reactivity. Int J Methods Psychiatr Res 2009 Mar;18(1):48-57 [FREE Full text] [doi: 10.1002/mpr.276] [Medline: 19195050]
30. Jones SE, Moore RC, Pinkham AE, Depp CA, Granholm E, Harvey PD. A cross-diagnostic study of Adherence to Ecological Momentary Assessment: comparisons across study length and daily survey frequency find that early adherence is a potent predictor of study-long adherence. Pers Med Psychiatry 2021 Nov;29-30:100085. [doi: 10.1016/j.pmip.2021.100085] [Medline: 34541425]
31. Swendsen J, Ben-Zeev D, Granholm E. Real-time electronic ambulatory monitoring of substance use and symptom expression in schizophrenia. Am J Psychiatry 2011 Feb;168(2):202-209. [doi: 10.1176/appi.ajp.2010.10030463] [Medline: 21078705]
32. Jones A, Remmerswaal D, Verveer I, Robinson E, Franken IH, Wen CK, et al. Compliance with ecological momentary assessment protocols in substance users: a meta-analysis. Addiction 2019 Apr 21;114(4):609-619 [FREE Full text] [doi: 10.1111/add.14503] [Medline: 30461120]
33. Williams MT, Lewthwaite H, Fraysse F, Gajewska A, Ignatavicius J, Ferrar K. Compliance with mobile ecological momentary assessment of self-reported health-related behaviors and psychological constructs in adults: systematic review and meta-analysis. J Med Internet Res 2021 Mar 03;23(3):e17023 [FREE Full text] [doi: 10.2196/17023] [Medline: 33656451]
34. Germine L, Reinecke K, Chaytor NS. Digital neuropsychology: challenges and opportunities at the intersection of science and software. Clin Neuropsychol 2019 Feb 06;33(2):271-286. [doi: 10.1080/13854046.2018.1535662] [Medline: 30614374]
35. The Role of Technology in Clinical Neuropsychology. Oxfordshire, England, UK: Oxford University Press; 2017.
36. Gershon RC, Cella D, Fox NA, Havlik RJ, Hendrie HC, Wagster MV. Assessment of neurological and behavioural function: the NIH Toolbox. Lancet Neurol 2010 Feb;9(2):138-139. [doi: 10.1016/S1474-4422(09)70335-7] [Medline: 20129161]
37. Weintraub S, Dikmen SS, Heaton RK, Tulsky DS, Zelazo PD, Bauer PJ, et al. Cognition assessment using the NIH Toolbox. Neurol 2013 Mar 11;80(Issue 11, Supplement 3):S54-S64. [doi: 10.1212/wnl.0b013e3182872ded]
38. Sheehan D, Lecrubier Y, Sheehan K, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20:22-33;quiz 34. [Medline: 9881538]
39. Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity and sensitivity. Br J Psychiatry 1978 Nov 29;133(5):429-435. [doi: 10.1192/bjp.133.5.429] [Medline: 728692]
40. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry 1979 Apr 29;134(4):382-389. [doi: 10.1192/bjp.134.4.382] [Medline: 444788]
41. Posner K, Brown GK, Stanley B, Brent DA, Yershova KV, Oquendo MA, et al. The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. Am J Psychiatry 2011 Dec;168(12):1266-1277 [FREE Full text] [doi: 10.1176/appi.ajp.2011.10111704] [Medline: 22193671]
42. Jeste DV, Palmer BW, Appelbaum PS, Golshan S, Glorioso D, Dunn LB, et al. A new brief instrument for assessing decisional capacity for clinical research. Arch Gen Psychiatry 2007 Aug 01;64(8):966-974. [doi: 10.1001/archpsyc.64.8.966] [Medline: 17679641]

43. Casaletto KB, Umlauf A, Beaumont J, Gershon R, Slotkin J, Akshoomoff N, et al. Demographically corrected normative standards for the English version of the NIH Toolbox cognition battery. J Int Neuropsychol Soc 2015 Jun 01;21(5):378-391. [doi: 10.1017/s1355617715000351]
44. Lomas J, Patel M, Moore R. NeuroUX: Mobile cognitive assessment games to measure 'cognition in the wild'. In: Proceedings of the CHIPLAY 2017. 2017 Presented at: CHIPLAY 2017; Oct 15, 2017; Amsterdam, Netherlands.
45. Moore RC, Paolillo EW, Sundermann EE, Campbell LM, Delgadillo J, Heaton A, et al. Validation of the mobile verbal learning test: illustration of its use for age and disease-related cognitive deficits. Int J Methods Psychiatr Res 2021 Mar 07;30(1):e1859 [FREE Full text] [doi: 10.1002/mpr.1859] [Medline: 33159488]
46. 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]
47. Desilver D. As it turns 6, a look at who uses the iPhone (no, not 'everybody'). Pew Research Center. 2013. URL: https://www.pewresearch.org/fact-tank/2013/06/29/as-it-turns-6-a-look-at-who-uses-the-iphone-no-not-everybody/ [accessed 2022-10-01]
48. Dupuy M, Misdrahi D, N'Kaoua B, Tessier A, Bouvard A, Schweitzer P, et al. Mobile cognitive testing in patients with schizophrenia: a controlled study of feasibility and validity. J de Thérapie Comportementale et Cognitive 2018 Dec;28(4):204-213. [doi: 10.1016/j.jtcc.2018.02.002]
49. Charalambous AP, Pye A, Yeung WK, Leroi I, Neil M, Thodi C, et al. Tools for app- and web-based self-testing of cognitive impairment: systematic search and evaluation. J Med Internet Res 2020 Jan 17;22(1):e14551 [FREE Full text] [doi: 10.2196/14551] [Medline: 31951218]
50. Bouvard A, Dupuy M, Schweitzer P, Revranche M, Fatseas M, Serre F, et al. Feasibility and validity of mobile cognitive testing in patients with substance use disorders and healthy controls. Am J Addict 2018 Oct 27;27(7):553-556. [doi: 10.1111/ajad.12804] [Medline: 30260085]
51. Depp CA, Kamarsu S, Filip TF, Parrish EM, Harvey PD, Granholm EL, et al. Ecological momentary facial emotion recognition in psychotic disorders. Psychol Med 2021 Jan 12:1-9. [doi: 10.1017/s0033291720004419]
52. Schweitzer P, Husky M, Allard M, Amieva H, Pérès K, Foubert-Samier A, et al. Feasibility and validity of mobile cognitive testing in the investigation of age-related cognitive decline. Int J Methods Psychiatr Res 2017 Sep 19;26(3):e1521 [FREE Full text] [doi: 10.1002/mpr.1521] [Medline: 27539327]
53. Iverson G, Karr J, Terry D, Garcia-Barrera M, Holdnack J, Ivins B, et al. Developing an executive functioning composite score for research and clinical trials. Arch Clin Neuropsychol 2020 Apr 20;35(3):312-325. [doi: 10.1093/arclin/acz070] [Medline: 31965141]
54. Siegal J. Study: iPads are the most responsive tablets in the world. BGR. URL: http://bgr.com/2013/10/09/tablet-touch-screen-responsiveness/2013 [accessed 2022-10-01]
55. Siegal J. Here's why typing on Android phones is harder than typing on an iPhone. BGR. URL: http://bgr.com/2013/09/20/iphone-android-touchscreen-responsiveness/2013 [accessed 2022-10-01]

Abbreviations

BD: bipolar disorder
C-SSRS: Columbia Suicide Severity Rating Scale
D-KEFS: Delis-Kaplan Executive Function System
EMA: ecological momentary assessment
EMCT: ecological momentary cognitive testing
NIH-TB: National Institutes of Health Toolbox
NIH-TB-CB: National Institutes of Health Toolbox Cognition Battery
VLMT: Variable Difficulty List Memory Test

Edited by R Kukafka, G Eysenbach; submitted 19.01.22; peer-reviewed by C Gould, P Harvey; comments to author 07.04.22; revised version received 25.04.22; accepted 29.05.22; published 29.07.22.

Please cite as:
Moore RC, Parrish EM, Van Patten R, Paolillo E, Filip TF, Bomyea J, Lomas D, Twamley EW, Eyler LT, Depp CA
Initial Psychometric Properties of 7 NeuroUX Remote Ecological Momentary Cognitive Tests Among People With Bipolar Disorder: Validation Study
J Med Internet Res 2022;24(7):e36665
URL: https://www.jmir.org/2022/7/e36665
doi:10.2196/36665
PMID:35904876

Original Paper

Item Response Theory Analyses of Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) Criteria Adapted to Screen Use Disorder: Exploratory Survey

Mathieu Boudard[1,2*], MD; Jean-Marc Alexandre[1,2*], MSc; Charlotte Kervran[1,3*], PhD; Louise Jakubiec[1,2], MD; Dvora Shmulewitz[4], PhD; Deborah Hasin[4], PhD; Lucie Fournet[1,2], MPhil; Christophe Rassis[1,2], MPhil; Patrice Claverie[5], MA; Fuschia Serre[1,2], PhD; Marc Auriacombe[1,2], MD

[1]Sanpsy (Sleep Addiction and Neuropsychiatry), CNRS (Centre National de la Recherche Scientifique), UMR (Unité Mixte de Recherche) 6033, University of Bordeaux, Bordeaux, France

[2]Pôle Interétablissement d'Addictologie, Centre Hospitalier Charles Perrens and Centre Hospitalier Universitaire de Bordeaux, Bordeaux, France

[3]MéRISP (Methods for population health intervention research), BPH (Bordeaux Population Health), INSERM (Institut National de la Santé et de la Recherche Médicale) Unité 1219, University of Bordeaux, Bordeaux, France

[4]Department of Psychiatry, Columbia University, New York, NY, United States

[5]Mairie de Martignas, Martignas, France

[*]these authors contributed equally

Corresponding Author:
Marc Auriacombe, MD
Sanpsy (Sleep Addiction and Neuropsychiatry), CNRS (Centre National de la Recherche Scientifique), UMR (Unité Mixte de Recherche) 6033
University of Bordeaux
146, rue Léo Saignat
Bordeaux, 33076
France
Phone: 33 607764371
Email: marc.auriacombe@u-bordeaux.fr

Abstract

Background: Screen use is part of daily life worldwide and morbidity related to excess use of screens has been reported. Some use of screens in excess could indicate a screen use disorder (ScUD). An integrative approach to ScUD could better fit the polymodal reality of screens, and concurrent problems with screens, than a split approach, activity by activity. In that paradigm, a pragmatic and operationalized approach to study a potential ScUD requires the use of common criteria, for all screens and activities done on screens, in a single questionnaire.

Objective: Our goals were (1) to describe screen uses in a general population sample and (2) to test the unidimensionality, local independence, and psychometric properties of the 9 Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) internet gaming disorder (IGD) criteria adapted to screen use in a community sample. We hypothesized that the 9 DSM-5 IGD criteria adapted to ScUD would show unidimensionality, local independence, and good discrimination, with criteria distributed on the severity continuum.

Methods: This cross-sectional survey in a French suburban city targeted adults and adolescents. A self-administered questionnaire covered the main types of screens used and their use for various activities in the past month. Presence of ScUD diagnostic criteria in past 12 months was also self-evaluated in the questionnaire. Factor and 2-parameter Item Response Theory analysis were used to investigate the dimensionality, local independence, and psychometric properties of the ScUD criteria.

Results: Among the 300 participants, 171 (57.0%) were female (mean age 27 years), 297 (99.0%) used screens, 134 (44.7%) reported at least one criterion (potential problem users), and 5 (1.7%) reported 5 or more criteria and endorsed an ScUD. The most endorsed criteria were loss of control (60/300, 20.0%) and preoccupation (52/300, 17.3%). Screen types used and screen activities differed between participants with no ScUD criteria and those with at least one ScUD criterion. The latter were more likely to have a computer as the most used screen type, and more video gaming, communication/social network, and watching news and research of information as activities. Unidimensionality was confirmed by all fit indices. Local independence was confirmed by the absence of residual correlation between the items. Criteria had relatively high factor loading, with loss of interest

in other recreational activities having the highest. However, criteria with the lowest factor loading all remained above the cut-offs, sanctioning unidimensionality. Most discriminating criteria were loss of interests, preoccupation, deceive/cover up, and risk/lose relationship/opportunities, which also provided the most information on the measurement of the latent trait.

Conclusions: We described screen uses in a French community sample and have shown that the adaptation of the DSM-5 IGD to "ScUD" has good psychometric validity and is discriminating, confirming our hypothesis. We suggest to use those criteria to assess potential "ScUD." Further studies should determine if all criteria are needed and whether others should be added.

(J Med Internet Res 2022;24(7):e31803) doi:10.2196/31803

KEYWORDS

screen media use; screen addiction; internet gaming disorder; screen use disorder; Item Response Theory

Introduction

Increased affordability and functionality of screen devices have contributed to making screen use part of current daily life worldwide [1-4]. Screen use facilitates communication for leisure-related activities (ie, video games, social media) and access to knowledge for education and work-related activities. However, some adverse consequences of using electronic screens have been reported. Sleep [5-7], visual problems [6,8], and overweight and obesity [9] have been associated with screen use. Excessive screen use has also been associated with a drop in academic accomplishments [10], psychiatric disorders [11], and suicide in adolescents [12]. All of these are related to duration of use and could be the expression of a potential addiction to screens [13,14]. Although the link and the direction of the link between screen use and increased mortality and morbidity remain to be confirmed [15,16], there is enough evidence to explore whether such a screen use disorder (ScUD) could be diagnosed for the purpose of prevention and treatment.

Based on clinical similarities with addictions, and the significant damages related to video game use, the American Psychiatric Association (APA) included internet gaming disorder (IGD) in the third section of the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5)* in expectation of further research [17,18]. The 9 IGD criteria were adapted from gambling disorder criteria, with a threshold of 5 to qualify for the diagnosis. Some criteria are common with those of substance use disorder. Differences are no craving or time spent criterion, an adverse negative mood, and a deceive/cover up criterion. Studies showed that IGD criteria have good psychometric validity with unidimensionality and good discrimination [19,20]. However, specific features of IGD are debated, including validity of the criteria and how to better operationally define them [18,21,22].

Many screen activities represent potentially addictive behaviors, and problematic media use has been studied on many screen types, such as "gaming disorder" [18,22,23], "smartphone use disorder" [24,25], and "internet addiction" [26]. Other authors have adapted IGD criteria to assess other potential behavioral addictions, such as "social media disorder" [27-29] and "screen media addiction" [14]. Considering clinical observations [30] and existing studies, we suggest combining these disorders into one "ScUD," characterized by the DSM-5 IGD criteria adapted to screen use [31]. We do not imply that screens are of themselves addictive, but that the combination of screen portability with ongoing internet access reduces time from decision to action and to positive reinforcement, which increases the addictive potential [32] of activities mediated by screen use. Screens offer a much higher availability, even permanent, of not just 1 activity but all of them at the same time, on the same medium, for almost everyone. Besides, internet connection may potentialize them (in terms of incitation, salience, rewards, problems, etc.). From a nosographic perspective, the study of a potential disorder of screen use with an integrative approach could better fit the polymodal reality of screens, and concurrent problems with screens, than a split approach, activity by activity. In that paradigm, a pragmatic and operationalized approach to study a potential ScUD requires the use of common criteria, for all screens and activities done on screens, in a single questionnaire.

Item Response Theory (IRT) postulates that a latent construct or trait that is not directly observable such as the proposed ScUD can be measured by a group of criteria [33]. These are the preferred analyses for assessing dimensional and structural validity of diagnostic criteria, such as IGD or substance use disorder criteria [19,20,34-37]. In recent studies on IGD that included gamers recruited via gaming websites or social media [19,20], screen media "addiction" in parents' reports of their children's behavior [14] showed that the IGD criteria fit well with the 1-factor model and that some criteria were more discriminant than others. However, to our knowledge, no study has yet assessed IGD criteria adapted to screen use using IRT among general population samples.

In 2015, Martignas-sur-Jalle (Nouvelle-Aquitaine, France) city council requested a local survey about screen uses (n=7400). This was an opportunity to conduct a general population survey of the IGD criteria adapted to screen use. Our goals were, in a suburban community sample, (1) to describe screen use and (2) to test the unidimensionality, local independence, and psychometric properties in terms of difficulty and discrimination of the 9 DSM-5 IGD criteria adapted to screen use. We hypothesized that the 9 DSM-5 IGD criteria adapted to ScUD would show unidimensionality, local independence, and good discrimination, with criteria distributed on the severity continuum.

Methods

Study Design

We designed an exploratory survey among the population of Martignas-sur-Jalle (n=7400). A task force with the University of Bordeaux, Charles Perrens Hospital Addiction Clinic,

Martignas-sur-Jalle city council, and population representatives was established to carry out and supervise the survey conducted from January 4, 2016, to February 25, 2016.

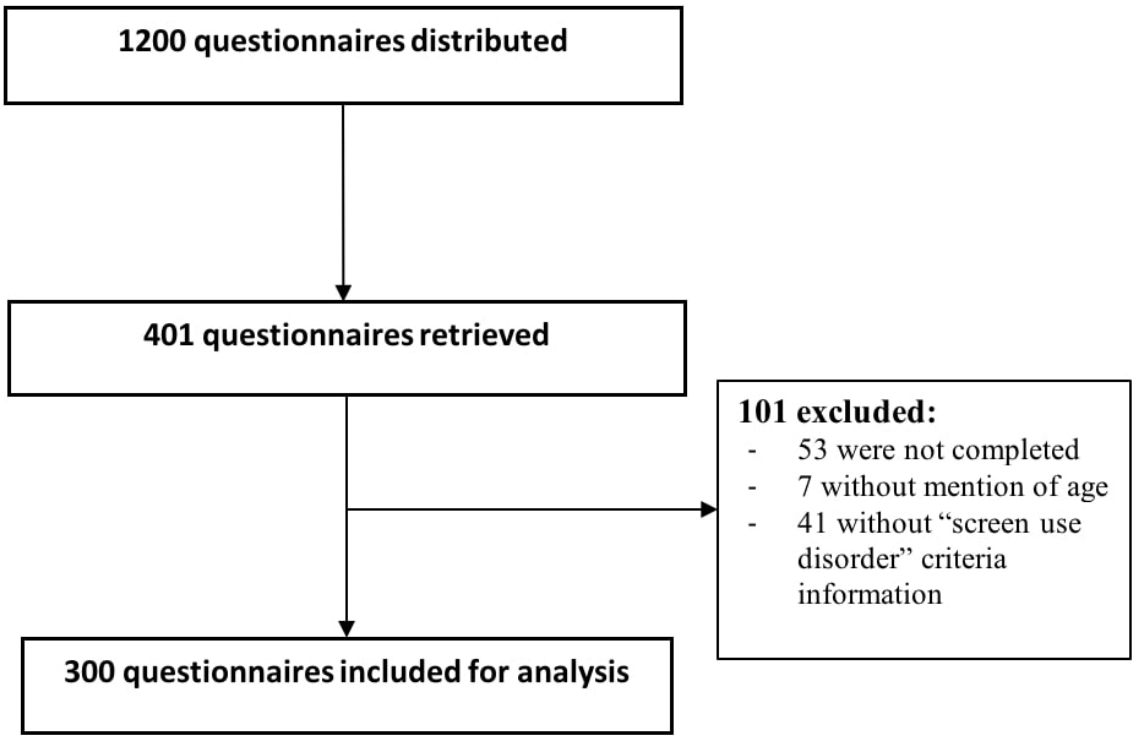
Participants

The study targeted all adults and adolescents from middle-school age (ie, from 11 to 12 years of age) with no upper age limitation. The task force agreed on this minimal age to assess screen users and ensure understanding of the questions. There were no exclusion criteria.

Procedure

Participants received the questionnaire from distribution points (all city services and schools) and returned them directly through ballot boxes or mail. Of the 1200 questionnaires distributed, 401 were returned. The response rate was 33.4% and the sample represented 6.6% of the target population of the city. After a quality check, 101 questionnaires were excluded (53 with no information, 7 without age, and 41 with ScUD questions not completed). The remaining 300 questionnaires were used for the database (Figure 1).

Figure 1. Flow chart of questionnaires selection process.



Survey Questionnaire

The survey questionnaire was a 10-15-minute 2-part self-administrated questionnaire including 49 closed-ended questions designed by the task force. The first part (37 questions) explored the main types of screens used over the past month (eg, TV, computers, smartphones, tablets, and handheld consoles) and for which activities (communication, social media, work, searching information on internet, other documentation, shopping, gaming, gambling, and others). The second part assessed each ScUD diagnostic criteria in the past 12 months (9 questions) and which screens and activities were considered problematic, that is, when at least one ScUD criterion was endorsed (1 question for screens and 1 for activities). We used the previously published French translation of the 9 IGD criteria [18] and adapted them to screen use (the term "videogames" was replaced by "screens"). The original French version and the translated version of the questionnaire are provided in Multimedia Appendix 1.

Measures

Our variables of interest were sociodemographic data (age, sex), screen use (converted into hours per day), activities, main screen used, main activities, prevalence of each diagnostic criteria, and ScUD. Activities were quantified by the number of days in the past 30 days (participants had to choose out of the following 4 options: every day or almost every day; more than 1 day out of 2; less than 1 day out of 2; and never or almost never). The main screen used was defined as the prevalence of participants for whom this screen was the most used (frequency over the past 30 days multiplied by the time per day). The main activity was defined for each activity as the prevalence of participants endorsing the activity on the main screen. ScUD was defined when 5 or more criteria were reported on the scale adapted from IGD (Table 1).

Table 1. Screen use disorder criteria.

Criteria	Internet gaming disorder	Screen use disorder
Preoccupation	Do you spend a lot of time thinking about games even when you are not playing, or planning when you can play next?	Do you spend a lot of time thinking about screens, even when you are not using them, or planning when you can use them next?
Withdrawal	Do you feel restless, irritable, moody, angry, anxious, or sad when attempting to cut down or stop gaming, or when you are unable to play?	Do you feel restless, irritable, moody, angry, anxious, or sad when attempting to cut down or stop using screens, or when you are unable to use screens?
Tolerance	Do you feel the need to play for increasing amounts of time, play more exciting games, or use more powerful equipment to get the same amount of excitement you used to get?	Do you feel the need to use screens for increasing amounts of time, use more exciting screens, or use more powerful equipment to get the same amount of excitement you used to get?
Loss of control	Do you feel that you should play less, but are unable to cut back on the amount of time you spend playing games?	Do you feel that you should use less screens, but are unable to cut back on the amount of time you spend using screens?
Loss of interest	Do you lose interest in or reduce participation in other recreational activities (hobbies, meetings with friends) due to gaming?	Do you lose interest in or reduce participation in other recreational activities (hobbies, meetings with friends) due to screens?
Continue despite problems	Do you continue to play games even though you are aware of negative consequences, such as not getting enough sleep, being late to school/work, spending too much money, having arguments with others, or neglecting important duties?	Do you continue to use screens even though you are aware of negative consequences, such as not getting enough sleep, being late to school/work, spending too much money, having arguments with others, or neglecting important duties?
Deceive/cover up	Do you lie to family, friends, or others about how much you game, or try to keep your family or friends from knowing how much you game?	Do you lie to family, friends, or others about how much you use screens, or try to keep your family or friends from knowing how much you use screens?
Escape adverse mood	Do you game to escape from or forget about personal problems, or to relieve uncomfortable feelings such as guilt, anxiety, helplessness, or depression?	Do you use screens to escape from or forget about personal problems, or to relieve uncomfortable feelings such as guilt, anxiety, helplessness, or depression?
Risk/lose relationship/opportunities	Do you risk or lose significant relationships, or job, educational, or career opportunities because of gaming?	Do you risk or lose significant relationships, or job, educational, or career opportunities because of screen use?

Statistical Analysis

Overview

We first described sociodemographic data. Quantitative variables were described by means and SD, and categorical variables with percentages. Adolescents and adults were analyzed together unless specified differently. Main activities and screen types for participants with no ScUD criteria versus those with at least one ScUD criteria were compared in univariate (Pearson tests) and multivariate analyses (logistic regression, controlled on age and gender). Statistical significance was set at $P<.05$. The prevalence of participants endorsing at least one ScUD criteria was compared between adults and teenagers. On an exploratory basis, participants with potential screen use problem (defined here as at least one criterion endorsed) were compared with those with no ScUD criterion.

Unidimensionality and Local Independence

To assess the dimensionality of the 9 criteria, a prerequisite to IRT, we fitted a 1-factor model using confirmatory factor analysis (CFA). Analysis was done using Mplus 8 [38]. Unidimensionality was confirmed when the CFA model showed adequate fit by comparative fit index or Tucker-Lewis Index of 0.95 or more and root mean squared error of approximation 0.06 or less [39]. Factor loadings below 0.40 were considered to be weakly related to the underlying construct [40].

We verified local independence between items using standardized z-scores with Mplus 8 [38,41]. Any significant residual correlation between the pairs of items (bivariate), after accounting for the underlying latent trait, would violate the assumption of local independence. Residual correlation between the items is observed if either the standardized *z*-scores for the different combinations of item responses are greater than 1.96 or below –1.96 (corresponding to a P value <.05), or if the chi-square value (an overall measure for both items, combining all the possible combinations) is greater than 3.84 ($P<.05$).

Item Response Theory

A 2-parameter logistic (2PL) IRT model was performed with the 9 criteria. Our scale was dichotomous and the 2PL model allowed us to examine the difficulty (inversely related to frequency; rarely endorsed criteria are considered more difficult) and discrimination (how well the criterion differentiated between respondents with high and low difficulty of the condition) of each criterion. Item characteristic curves (ICCs) were generated to display the estimated probability of endorsing each criterion across the underlying continuum. In the ICC, the difficulty parameter was the point on the x-axis where the probability of endorsing a criterion was 0.5 (curve toward the right indicates criteria of greater difficulty), and discrimination is the slope of the curve at that point (steeper slopes indicate greater discrimination). We generated item information curves, an indicator on how each item contributes variably to the total test

information. Total information curves were generated to show their ability to discriminate individuals along the latent trait severity spectrum [33,40,42].

Description of the sample (mean, SD, and percentage) was performed with JMP; CFA and IRT (psychometric analysis) were performed with Mplus 8 [38].

Ethics Approval

The survey was anonymous and confidential, and met French regulation ethics standards for noninterventional research after institutional review board (Sanpsy/University of Bordeaux) review [43]. Participation was voluntary with no financial compensation. The questionnaire was distributed with an information note presenting the investigation, consent collection, confidentiality, and legal issues.

Results

Sociodemographic Information

Of the 300 participants, 171 were women (57.0%), mean age was 27 years (SD 18.9 years), and 160 were under 18 years (53.3%). The youngest participant was 11 years and the oldest was 84 years. Almost all participants (n=297, 99.0%) reported daily screen use (Table 2).

Table 2. Demographic characteristics, screen use (any), and screen use disorder (n=300).

Characteristics	Sample
Age, mean (SD)	27 (18.9)
Age, median	15
Males, n (%)	129 (43.0)
<18-year olds, n (%)	160 (53.3)
Screen use (every day), n (%)	297 (99.0)
Screen use disorder criteria (cumulative), n (%)	
0	166 (55.3)
≥1	134 (44.7)
≥2	58 (19.3)
≥3	23 (7.7)
≥4	7 (2.3)
≥5	5 (1.7)
≥6	3 (1.0)
7	1 (0.3)

Screen Use Disorder Diagnosis

Most of the sample reported no criteria (n=166, 55.3%), 134 participants (44.7%) reported at least one criterion (potential problem users), and 5 participants (1.7%) reported 5 criteria or more and qualified for a potential ScUD (Table 2). Adolescents (defined as 11-17 years; mean age 12.92 years, SD 1.50 years) were significantly more likely to endorse at least one ScUD criteria than adults (defined as being aged above 18 years, mean age 43.2 years, SD 16.5 years; 97/300, 32.3% vs 37/300, 12.3%; $P<.001$).

The prevalence of each criterion is reported in Table 3. The most endorsed were *loss of control* (60/300, 20.0%) and *preoccupation* (52/300, 17.3%). The less endorsed were *losing an opportunity* (6/300, 2.0%) and *tolerance* (7/300, 2.3%).

Table 3. Parameter estimates from confirmatory factor analysis/Item Response Theory analysis in screen use disorder.

Screen use disorder criteria	Factor loading[a]	Screen use (n=300)				
		Item Response Theory parameters				
	1-factor model	Prevalence (N=300), n (%)	*(b)* Difficulty (SE)	Difficulty rank	*(a)* Discrimination (SE)	*(c)* Discrimination rank
Preoccupation	0.726	52 (17.3)	1.279 (0.224)	1	1.882 (0.618)	2
Withdrawal	0.457	10 (3.3)	3.656 (1.515)	9	1.058 (0.569)	7
Tolerance	0.493	7 (2.3)	3.290 (0.855)	8	1.404 (0.500)	5
Loss of control	0.477	60 (20.0)	1.806 (0.484)	2	0.884 (0.284)	9
Loss of interests	0.779	21 (7.0)	1.962 (0.350)	3	2.027 (0.714)	1
Continue despite problems	0.499	44 (14.7)	2.009 (0.468)	4	1.047 (0.317)	8
Deceive/cover up	0.649	10 (3.3)	2.658 (0.523)	5	1.735 (0.564)	3
Escape adverse mood	0.568	21 (7.0)	2.664 (0.704)	6	1.174 (0.429)	6
Risk/lose relationship/opportunities	0.650	6 (2.0)	3.020 (0.856)	7	1.721 (0.823)	4

[a]Model fit indices: comparative fit index 1.000; Tucker-Lewis Index 1.026; root mean square error of approximation ≤0.0001.

Dimensionality, Local Independence, and IRT Analysis

Unidimensionality was confirmed by all fit indices (comparative fit index 1.000; Tucker-Lewis Index 1.026; root mean square error of approximation ≤0.0001; and factor loading ≥0.4 for each criterion). Local independence was confirmed by the absence of residual correlation between the items (minimum and maximum standardized *z*-scores for the different combinations of item responses were equal to –1.042 and 1.129, respectively; maximal chi-square value was 2.008). All criteria had relatively high factor loading except *tolerance* (0.493), *withdrawal* (0.457), and *loss of control* (0.477), but these also remained above the cut-offs sanctioning unidimensionality. Factor loading for *loss of interest* (0.779) was higher than for any other diagnostic criterion, followed by *preoccupation* (0.726). The criterion *preoccupation* (1.279) had the lowest difficulty to be endorsed, followed by *loss of control* and *loss of interest*. Inversely, the *withdrawal* and *tolerance* criteria showed the highest difficulty. Discrimination parameters ranged from 0.884 to 2.027, indicating a good ability to delineate individuals who were higher versus lower to the latent trait (ICC; Figure 2). Both *Loss of interest* (2.027) and *preoccupation* (1.279) criteria showed a higher discrimination, while *loss of control* showed a lower discrimination (0.884) compared with other criteria (Table 3).

Item information curves (Figure 3) showed that most discriminating criteria were, in order, *loss of interests*, *preoccupation*, *deceive/cover up*, and *risk/lose relationship/opportunities*, which also provided the most information on the measurement of the latent trait. *Loss of interests* and *preoccupation* criteria also provided the greatest amount of information and high precision across the latent trait severity continuum of ScUD. Loss of control criterion was identified as the least discriminating and the least informative.

Total information curves (Figure 3) showed an increased information across the severity spectrum for the 9 IGD criteria group. Removing the *loss of control* criterion did not seem to affect the ability of the test to capture the disorder phenomenon. However, removing the *loss of interests* criterion changed the amount of severity information provided by the test. The 3 models brought roughly the same range of severity.

Figure 2. ICC for IGD criteria adapted to screen use disorder in the general population sample of a French suburban city. ICC: item characteristics curve; IGD: internet gaming disorder.

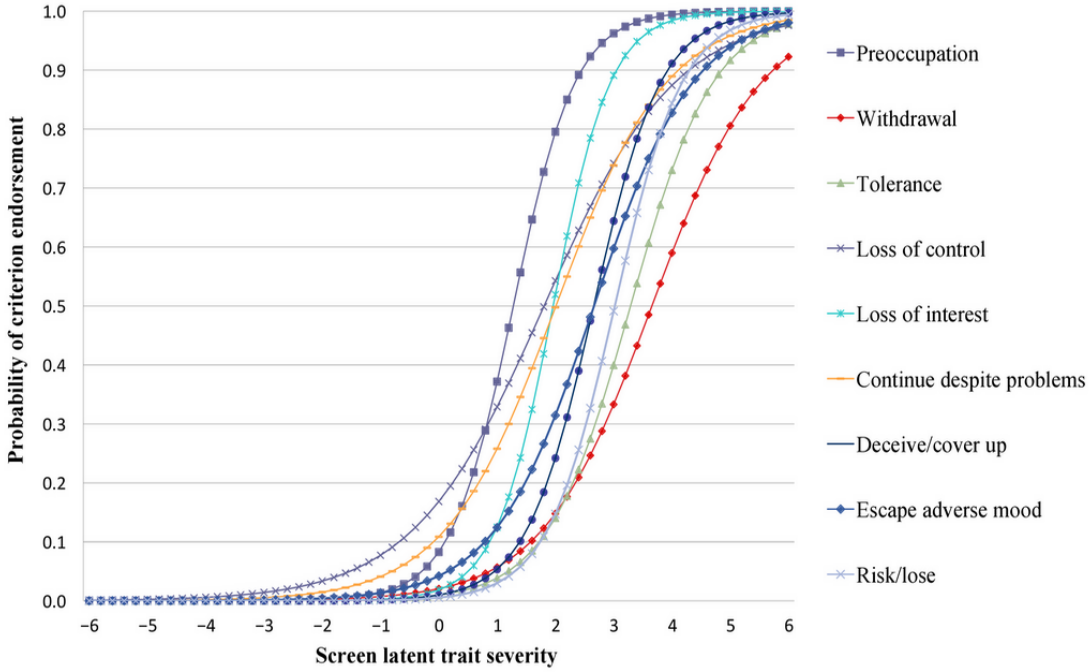
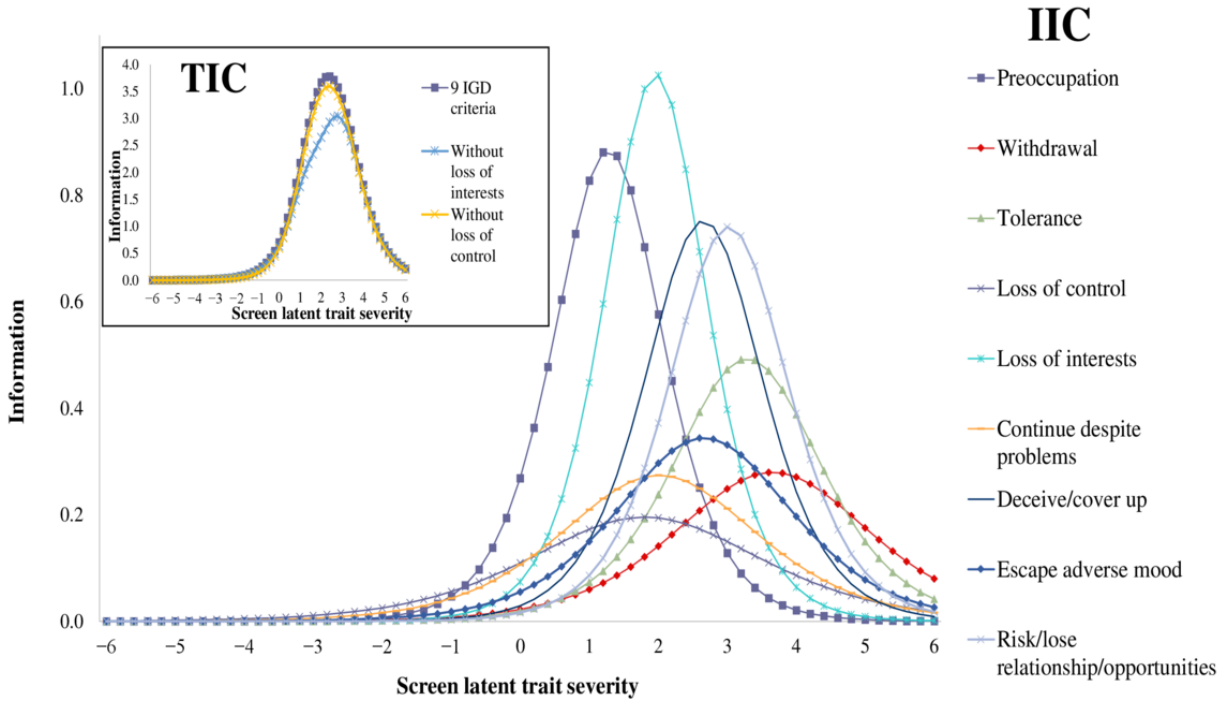


Figure 3. IICs and TICS for IGD criteria adapted to screen use disorder in the general population sample of a French suburban city. IGD: internet gaming disorder; IIC: item information curve; TIC: total information curve.



Screen Use and Screen Activities

In univariate analysis, participants with no ScUD criterion were more likely to report *television* (P<.001) as the most used screen (Table 4). Participants with at least one ScUD criterion were more likely to have *smartphone* (P=.04) and *computer* (P=.04) as the most used screens, which they also reported as the most problematic screens: *smartphone* (69/269, 25.7%) and *computer* (61/269, 22.7%). For *tablets* and *handled console* there was no

difference between participants with and without at least one ScUD criteria.

Compared with participants with no ScUD criteria, participants with at least one ScUD criteria reported more *video gaming* (*P*<.001) and *communication/social network* (*P*<.001), which they also reported as the most problematic activity: 36.6% (71/194) and 31.4% (61/194), respectively.

In multivariate analysis, when controlled on age and gender, participants with at least one ScUD criterion were more likely to have a *computer* (*P*=.004) as the most used screen type. For activities, they reported more *videogaming* (*P*=.002) and *communication/social network* (*P*=.03) compared with participants with no ScUD criteria. Besides, a new association was found between having at least one ScUD criterion and *watching news and research of information* (*P*=.002) that was not observed in the univariate analysis.

Table 4. Main screen and activity for participants with no ScUD[a] criteria and at least one ScUD criteria. Description of screen type and activities considered as problematic for participants with at least one ScUD criteria.

Activities (several answers possible)	Participants with no ScUD criteria (n=166)	Participants with 1 or more ScUD criteria (n=134)	Univariate analysis: *P* value (Pearson)	Multivariate analysis: adjusted *P* value (logistic regression)
Screen type (several answers possible), n (%)				
TV	103 (62.0)	57 (42.5)	<.001	.06[b]
Smartphone	55 (33.1)	60 (44.8)	.04	.41
Computer	24 (14.5)	32 (23.9)	.04	.004
Tablet	20 (12.0)	23 (17.2)	.21	.73
Handheld console	7 (4.2)	11 (8.2)	.15	.30
Screen type reported as problematic (several answers possible; n=269), n (%)				
TV	—[c]	58 (21.6)	—	—
Smartphone	—	69 (25.7)	—	—
Computer	—	61 (22.7)	—	—
Tablet	—	45 (16.7)	—	—
Handheld console	—	31 (11.5)	—	—
Other	—	5 (1.9)	—	—
Screen activities reported as problematic (several answers possible; n=194)				
News and information	—	17 (8.8)	—	—
Work-related activities	—	8 (4.1)	—	—
Others	—	27 (13.9)	—	—
Communication/social	—	71 (36.6)	—	—
Video gaming	—	61 (31.4)	—	—
Purchase	—	6 (3.1)	—	—
Gambling	—	4 (2.1)	—	—

[a]ScUD: screen use disorder.

[b]Not significant.

[c]Not applicable.

Discussion

Principal Findings

This is the first study to combine description of screen use and exploration of the dimensionality and psychometric validity of the 9 IGD DSM-5 criteria adapted to a potential "ScUD" among a general population sample. Prevalence of ScUD was 1.7% (5/300) in our sample. Our results confirm the initial hypothesis of unidimensionality of the 9 IGD DSM-5 criteria adapted to ScUD.

Almost all participants (297/300, 99.0%) of this survey used screens daily, reflecting a high level of equipment use in daily life. ScUD criteria were characterized by the DSM-5 IGD criteria adapted to screen use. The majority of our participants (166/300, 55.3%) self-reported none of the criteria in the past 12 months. However, a notable proportion (134/300, 44.7%) self-reported at least one criterion and a screen type or screen activity as problematic in the past 12 months. This can be interpreted as a need for support and advice for better use of screens in that population. "Screen addiction" prevalence (≥5 criteria endorsed) was 1.7% (5/300), which is in range with the

prevalence of IGD (2.0%) in population-based studies [23,44,45]. Two recent meta-analysis on gaming disorder prevalence, a "screen-related addiction," showed prevalence in the same order of magnitude, 3.0% and 3.3%, respectively [44,45]. Our results are interesting in that they go against lay beliefs of a very high prevalence of "screen addiction." For those people satisfying 5 or more ScUD criteria, a persistent and recurrent use of screens leading to clinically significant impairment or distress could be assumed, similar to IGD in the DSM-5 [17,18]. In this particular situation, it should be assumed that advice on screen use would be insufficient, and that an addiction-oriented intervention would be useful. There were more adolescents than adults with at least one ScUD criterion (97/300, 32.3% vs 37/300, 12.3%). As many as 2 adults and 3 teenagers met the threshold of 5 criteria for ScUD.

Screens most used differed between participants with at least one ScUD criteria or no criteria. When controlled for age and gender, participants with at least one ScUD criterion were significantly more likely to use computers as the main screen. This may be explained by the activities performed on computers. These participants reported more *video gaming*, *communication/social network*, and *watching news and research of information*, all of which are commonly done on computers. Screens and activities reported as problematic by participants with at least one ScUD criterion were similar to the screens used (eg, smartphone, computers) and activities (eg, video gaming, communication/social network) performed the most, a result that may be of interest for prevention. There was a group of screen users that reported some problem with use and as such is likely to be responsive to interventions focused on related support.

Our study showed unidimensionality of the 9 IGD DSM-5 criteria adapted to ScUD. The model showed adequate fit and the criteria reflected 1 underlying latent trait (ScUD). Moreover, we found no residual correlation between the items, and thus confirmed local independence, a fundamental assumption in IRT models. This means that the items were correlated only through the latent trait that the test is measuring [46]. Some criteria had specific psychometric characteristics. *Loss of interest* (losing interest or reducing participation in other recreational activities) and *preoccupation* (being absorbed by screen use and thinking about it) loaded more strongly than other diagnostic criteria, indicating that they fit well with the 1-factor model, similar to results from a parent-reported survey of screen media "addiction" in children [14]. These criteria were among the more frequently endorsed, and had higher discrimination than others. Thus, both *loss of interest* and *preoccupation* criteria seem to capture the less severe end of the diagnostic spectrum, and the criteria well differentiated between respondents with high and low screen use severity. By identifying participants with less severe ScUD, these items are potentially useful as early indicators of ScUD [47]. It would be interesting to assess, within a prospective cohort of adolescents, whether the occurrence of these criteria predicts a subsequent ScUD.

Withdrawal and *tolerance* criteria had the lowest factor loading and showed the highest difficulty and moderate discrimination power, similar to results in a general population study of children [14]. Our results suggest that these criteria may not be relevant to define ScUD. By contrast, in some IGD surveys including population of video gamers with significant gaming time, *withdrawal* and *tolerance* had higher factor loadings and seemed discriminating [19,20], suggesting that very high and regular level of gaming practice may promote tolerance and withdrawal symptoms. Recently, the World Health Organization (WHO) specified its own gaming disorder criteria in the 11th revision of the International Classification of Diseases (ICD-11) [48]. *Tolerance* and *withdrawal* criteria were removed, as well as *preoccupation*, *deceive/cover up*, and *escape adverse mood*. Additional studies among the general population are thus needed to determine to what extent withdrawal and tolerance are related to the intensity of screen use and characterize potential ScUD.

The *loss of control* criterion (feeling that you should use less screens, but being unable to cut back on the amount of time spent watching it) had a lower factor loading, a lower discrimination power, and was among the less difficult (more frequent) criteria. This suggests that this criterion is frequent in a population without ScUD, perhaps due to high overall screen use exposure [1,3]. Including a criterion with poor discrimination may increase the risk for false-positive diagnosis, especially at the lower range of difficulty (high frequency) [47]. In previous studies on IGD [19,20] this criterion had low standard in terms of factor loading, discrimination, and difficulty. However, this result is questionable because this criterion is reported to be a central criterion of addiction [17,49]. By contrast, in another study about screen addiction, *loss of control* showed the highest factor loading in children [14], possibly because reports were from children's caregivers, and cessation of use is a source of conflict between parents and children. More studies are therefore needed to evaluate the potential importance of this criterion in ScUD.

Limitations

Study limitations are to be noted. This was a convenience sample with a somewhat low response rate. Survey respondents represented 6.60% (401/6075) of the target population (men and women above 11 years from Martignas-sur-Jalle). Compared with the target population, our final sample was younger (24 years vs 40.5 years), mainly due to a higher proportion of 12-18 year olds (160/300, 53.3%, in our sample vs 784/6075, 12.9%, in the target population). Gender ratio was comparable (3159/6075, 52.0%, women in target population vs 171/300, 57.0%, in our sample). As our questionnaire was self-administered, risk of participant misinterpretation should be considered. However, we used the operationalized formulations for IGD assessment [18]. As a result of missing information, 101 questionnaires (responses) were excluded and there was a higher rate of adolescents among our sample. This could suggest that the questionnaire may have been of little interest to some participants, or might have been difficult to understand, or that adolescents might feel more concerned by this survey. An important element for the validity of the ScUD diagnostic criteria is to determine whether criteria or criteria sets function differently across population subgroups, such as age or sex. As our sample is composed of adolescents and adults, it would be interesting to see whether criteria behave differently according to age. However, in this study, the prevalence of some criteria was too small and thus such an analysis could not be

performed here. Additional data in samples more likely to endorse ScUD criteria should be collected and analyzed for differential item functioning. Further studies should investigate the relationships between ScUD items to determine whether there is some local dependence, indicating a possible redundancy. Finally, because craving was not part of IGD criteria [18], no craving criterion was assessed. Some studies suggest that craving should be included [50,51], as it has a high prevalence in samples of those with IGD [52,53], and is the most specific criterion for many substance use disorders [37]. Additional studies should thus be carried out by including craving.

Conclusions

We described screen use in a French community sample and have shown that the adaptation of the DSM-5 IGD criteria to "ScUD" has good psychometric validity. Endorsement of diagnostic criteria in the past 12 months could be interpreted as current complaints and impairment of the users, strengthening the possibility for ScUD to qualify as a disorder. Further studies are needed to confirm the validity of ScUD diagnosis and its negative consequences. We suggest that there may be similarities between different screen-related addictions, thus allowing for a broader tool to encompass the screen activities. Future studies will have to determine whether the type of screen/screen activity is related to the likelihood of ScUD diagnosis, the validity of a craving criterion, if all criteria are needed or if some should be removed or replaced, and if the diagnosis threshold of 5 is appropriate. Screen use and its consequences represent an important emerging field for addiction research.

Acknowledgments

The authors acknowledge all participants for their contribution and are grateful to Martignas city representatives for their collaboration. Since the study was finished some contributors have changed affiliation. LJ is now with the Institut de Neurosciences Cognitives et Intégratives d'Aquitaine (INCIA) at University of Bordeaux. CK is now with Methods for Population Health Intervention Research (MéRISP), Institut National de la Santé et de la Recherche Médicale (Inserm) Unit 1219, Bordeaux Population Health Research Center (BPH) at University of Bordeaux. PC is with the city of Saint-Médard-en-Jalles.

Authors' Contributions

MA was the overall principal investigator of the study. J-MA, LJ, MB, CR, PC, FS, and MA developed study protocol and the questionnaire. MB, J-MA, and CK performed analysis. MB wrote the first draft of the manuscript, which was edited by J-MA, LJ, CK, DS, DH, and MA. PC coordinated Martignas city representatives in contact with CR, and monitored questionnaire dissemination and collection. FS and LF provided methodological support. All authors contributed to and have approved the final manuscript.

Conflicts of Interest

None declared

Multimedia Appendix 1
English and French versions of the questionnaire.
[PDF File (Adobe PDF File), 192 KB - jmir_v24i7e31803_app1.pdf]

References

1. Observatoire Français des Drogues et des Toxicomanies (OFDT). Niveaux d'usages des ecrans à la fin de l'adolescence en 2017. OFDT. 2019. URL: https://www.ofdt.fr/BDD/publications/docs/eisxapza.pdf [accessed 2022-06-22]
2. United States Census Bureau. Computer and Internet Use in the United States: 2015. United States Census Bureau. 2017. URL: https://www.census.gov/content/dam/Census/library/publications/2017/acs/acs-37.pdf [accessed 2022-06-22]
3. Ofcom. Children and parents: Media Use and Attitudes Report. Ofcom. 2020. URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0023/190616/children-media-use-attitudes-2019-report.pdf [accessed 2022-06-22]
4. Screen Time: UK vs US vs The Rest Of The World Compared. Businessfibre. 2019. URL: https://businessfibre.co.uk/screen-time/ [accessed 2022-06-22]
5. Reynolds AC, Meltzer LJ, Dorrian J, Centofanti SA, Biggs SN. Impact of high-frequency email and instant messaging (E/IM) interactions during the hour before bed on self-reported sleep duration and sufficiency in female Australian children and adolescents. Sleep Health 2019 Feb;5(1):64-67. [doi: 10.1016/j.sleh.2018.10.008] [Medline: 30670168]
6. Benchebra L, Alexandre J, Dubernet J, Fatséas M, Auriacombe M. [Gambling and Gaming disorders and physical health of players: A critical review of the literature]. Presse Med 2019 Dec;48(12):1551-1568. [doi: 10.1016/j.lpm.2019.10.014] [Medline: 31767247]
7. Carter B, Rees P, Hale L, Bhattacharjee D, Paradkar MS. Association between portable screen-based media device access or use and sleep outcomes: a systematic review and meta-analysis. JAMA Pediatr 2016 Dec 01;170(12):1202-1208 [FREE Full text] [doi: 10.1001/jamapediatrics.2016.2341] [Medline: 27802500]

8. Kozeis N. Impact of computer use on children's vision. Hippokratia 2009 Oct;13(4):230-231 [FREE Full text] [Medline: 20011087]
9. Tremblay MS, LeBlanc AG, Kho ME, Saunders TJ, Larouche R, Colley RC, et al. Systematic review of sedentary behaviour and health indicators in school-aged children and youth. Int J Behav Nutr Phys Act 2011 Sep 21;8:98 [FREE Full text] [doi: 10.1186/1479-5868-8-98] [Medline: 21936895]
10. Adelantado-Renau M, Moliner-Urdiales D, Cavero-Redondo I, Beltran-Valls MR, Martínez-Vizcaíno V, Álvarez-Bueno C. Association between screen media use and academic performance among children and adolescents: a systematic review and meta-analysis. JAMA Pediatr 2019 Nov 01;173(11):1058-1067 [FREE Full text] [doi: 10.1001/jamapediatrics.2019.3176] [Medline: 31545344]
11. Boers E, Afzali MH, Conrod P. Association of media screen time use with depression in adolescents-reply. JAMA Pediatr 2020 Feb 01;174(2):210. [doi: 10.1001/jamapediatrics.2019.4920] [Medline: 31860010]
12. Twenge JM, Joiner TE, Rogers ML, Martin GN. Increases in depressive symptoms, suicide-related outcomes, and suicide rates among U.S. adolescents after 2010 and links to increased new media screen time. Clinical Psychological Science 2017 Nov 14;6(1):3-17. [doi: 10.1177/2167702617723376]
13. Balhara YPS, Verma K, Bhargava R. Screen time and screen addiction: Beyond gaming, social media and pornography-A case report. Asian J Psychiatr 2018 Jun;35:77-78. [doi: 10.1016/j.ajp.2018.05.020] [Medline: 29803121]
14. Domoff SE, Harrison K, Gearhardt AN, Gentile DA, Lumeng JC, Miller AL. Development and validation of the problematic media use measure: a parent report measure of screen media "addiction" in children. Psychol Pop Media Cult 2019 Jan;8(1):2-11 [FREE Full text] [doi: 10.1037/ppm0000163] [Medline: 30873299]
15. Twenge JM. Why increases in adolescent depression may be linked to the technological environment. Curr Opin Psychol 2020 Apr;32:89-94. [doi: 10.1016/j.copsyc.2019.06.036] [Medline: 31415993]
16. Sedgwick R, Epstein S, Dutta R, Ougrin D. Social media, internet use and suicide attempts in adolescents. Curr Opin Psychiatry 2019 Nov;32(6):534-541. [doi: 10.1097/YCO.0000000000000547] [Medline: 31306245]
17. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (5th ed). Washington, DC: American Psychiatric Association; 2013.
18. 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]
19. Király O, Sleczka P, Pontes HM, Urbán R, Griffiths MD, Demetrovics Z. Validation of the Ten-Item Internet Gaming Disorder Test (IGDT-10) and evaluation of the nine DSM-5 Internet Gaming Disorder criteria. Addict Behav 2017 Jan;64:253-260. [doi: 10.1016/j.addbeh.2015.11.005] [Medline: 26632194]
20. Schivinski B, Brzozowska-Woś M, Buchanan EM, Griffiths MD, Pontes HM. Psychometric assessment of the Internet Gaming Disorder diagnostic criteria: An Item Response Theory study. Addict Behav Rep 2018 Dec;8:176-184 [FREE Full text] [doi: 10.1016/j.abrep.2018.06.004] [Medline: 30505924]
21. Auriacombe M, Denis C, Fatséas M. Commentary on Scharkow et al. (2014): Welcomed and yet to come. Are we looking at what we are looking for? Addiction 2014 Nov;109(11):1918-1919. [doi: 10.1111/add.12723] [Medline: 25297961]
22. Leouzon H, Alexandre J, Fatséas M, Auriacombe M. L'addiction aux jeux vidéo dans le DSM-5, controverses et réponses relatives à son diagnostic et sa définition. Annales Médico-psychologiques, revue psychiatrique 2019 Sep;177(7):610-623. [doi: 10.1016/j.amp.2019.03.013]
23. Paulus FW, Ohmann S, von Gontard A, Popow C. Internet gaming disorder in children and adolescents: a systematic review. Dev Med Child Neurol 2018 Jul 06;60(7):645-659 [FREE Full text] [doi: 10.1111/dmcn.13754] [Medline: 29633243]
24. Yu S, Sussman S. Does smartphone addiction fall on a continuum of addictive behaviors? Int J Environ Res Public Health 2020 Jan 08;17(2):31936316 [FREE Full text] [doi: 10.3390/ijerph17020422] [Medline: 31936316]
25. Sohn S, Rees P, Wildridge B, Kalk NJ, Carter B. Prevalence of problematic smartphone usage and associated mental health outcomes amongst children and young people: a systematic review, meta-analysis and GRADE of the evidence. BMC Psychiatry 2019 Nov 29;19(1):356 [FREE Full text] [doi: 10.1186/s12888-019-2350-x] [Medline: 31779637]
26. Cheng C, Li AY. Internet addiction prevalence and quality of (real) life: a meta-analysis of 31 nations across seven world regions. Cyberpsychol Behav Soc Netw 2014 Dec;17(12):755-760 [FREE Full text] [doi: 10.1089/cyber.2014.0317] [Medline: 25489876]
27. van den Eijnden RJ, Lemmens JS, Valkenburg PM. The Social Media Disorder scale. Computers in Human Behavior 2016 Aug;61:478-487. [doi: 10.1016/j.chb.2016.03.038]
28. Fung S. Cross-cultural validation of the Social Media Disorder scale. Psychol Res Behav Manag 2019;12:683-690 [FREE Full text] [doi: 10.2147/PRBM.S216788] [Medline: 31695527]
29. Chen I, Strong C, Lin Y, Tsai M, Leung H, Lin C, et al. Time invariance of three ultra-brief internet-related instruments: Smartphone Application-Based Addiction Scale (SABAS), Bergen Social Media Addiction Scale (BSMAS), and the nine-item Internet Gaming Disorder Scale- Short Form (IGDS-SF9) (Study Part B). Addict Behav 2020 Feb;101:105960. [doi: 10.1016/j.addbeh.2019.04.018] [Medline: 31072648]
30. Sharma M, Sharma MK, Anand N, John N, Sahu M, Chakraborty Thakur P, et al. Binge watching: An emerging manifestation of technology use. Asian J Psychiatr 2019 Oct;45:81-82. [doi: 10.1016/j.ajp.2019.08.009] [Medline: 31526919]

31. Alexandre J, Rassis C, Serre F, Fatseas M, Auriacombe M. A survey of potential Internet Gaming Disorder extended to screen use in a community sample. SANPSY. 2017. URL: https://sanpsy.u-bordeaux.fr/fr/categorie2/news/page11 [accessed 2022-06-22]
32. Ainslie G. The Picoeconomics of Addiction. In: Pickard H, Ahmed SH, editors. The Routledge Handbook of the Philosophy and Science of Addiction. London, UK: Routledge; Jun 11, 2018:34-44.
33. Embretson SE, Reise SP. Item response theory for psychologists. Mahwah, NJ: Lauwrence Erlbaum Associates Publishers; 2000:371.
34. Shmulewitz D, Keyes K, Beseler C, Aharonovich E, Aivadyan C, Spivak B, et al. The dimensionality of alcohol use disorders: results from Israel. Drug Alcohol Depend 2010 Sep 01;111(1-2):146-154 [FREE Full text] [doi: 10.1016/j.drugalcdep.2010.04.002] [Medline: 20537809]
35. Saha TD, Compton WM, Chou SP, Smith S, Ruan WJ, Huang B, et al. Analyses related to the development of DSM-5 criteria for substance use related disorders: 1. Toward amphetamine, cocaine and prescription drug use disorder continua using Item Response Theory. Drug Alcohol Depend 2012 Apr 01;122(1-2):38-46 [FREE Full text] [doi: 10.1016/j.drugalcdep.2011.09.004] [Medline: 21963414]
36. Mewton L, Slade T, McBride O, Grove R, Teesson M. An evaluation of the proposed DSM-5 alcohol use disorder criteria using Australian national data. Addiction 2011 May;106(5):941-950. [doi: 10.1111/j.1360-0443.2010.03340.x] [Medline: 21205055]
37. Kervran C, Shmulewitz D, Serre F, Stohl M, Denis C, Hasin D, et al. Item Response Theory analyses of DSM-5 substance use disorder criteria in French outpatient addiction clinic participants. How much is craving special? Drug Alcohol Depend 2020 Jul 01;212:108036. [doi: 10.1016/j.drugalcdep.2020.108036] [Medline: 32464467]
38. Muthén LK, Muthén BO. Mplus User's Guide. Eighth Edition. Muthén & Muthén. Los Angeles, CA: Muthén & Muthén; 1998. URL: https://www.statmodel.com/download/usersguide/MplusUserGuideVer_8.pdf [accessed 2022-06-22]
39. 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]
40. Shmulewitz D, Keyes K, Wall MM, Aharonovich E, Aivadyan C, Greenstein E, et al. Nicotine dependence, abuse and craving: dimensionality in an Israeli sample. Addiction 2011 Sep;106(9):1675-1686 [FREE Full text] [doi: 10.1111/j.1360-0443.2011.03484.x] [Medline: 21545668]
41. Muthén LK, Muthén BO. Mplus Documentation. Mplus. URL: https://www.statmodel.com/ [accessed 2022-06-19]
42. Patel ZS, Jensen-Doss A, Zopluoglu C. Illustrating the applicability of IRT to implementation science: examining an instrument of therapist attitudes. Adm Policy Ment Health 2021 Sep;48(5):921-935. [doi: 10.1007/s10488-021-01139-1] [Medline: 33929639]
43. Commission Nationale de l'Informatique et des Libertés (CNIL). Méthodologie de référence MR-003. Délibération n°2018-154 du 3 mai 2018. French Government. 2018. URL: https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000037187443 [accessed 2022-06-22]
44. Kim HS, Son G, Roh E, Ahn W, Kim J, Shin S, et al. Prevalence of gaming disorder: A meta-analysis. Addict Behav 2022 Mar;126:107183. [doi: 10.1016/j.addbeh.2021.107183] [Medline: 34864436]
45. Stevens MW, Dorstyn D, Delfabbro PH, King DL. Global prevalence of gaming disorder: A systematic review and meta-analysis. Aust N Z J Psychiatry 2021 Jun 07;55(6):553-568. [doi: 10.1177/0004867420962851] [Medline: 33028074]
46. Christensen KB, Makransky G, Horton M. Critical values for Yen's : Identification of local dependence in the Rasch Model using residual correlations. Appl Psychol Meas 2017 May 16;41(3):178-194 [FREE Full text] [doi: 10.1177/0146621616677520] [Medline: 29881087]
47. Chung T, Martin CS, Maisto SA, Cornelius JR, Clark DB. Greater prevalence of proposed DSM-5 nicotine use disorder compared to DSM-IV nicotine dependence in treated adolescents and young adults. Addiction 2012 Apr;107(4):810-818 [FREE Full text] [doi: 10.1111/j.1360-0443.2011.03722.x] [Medline: 22092543]
48. World Health Organisation (WHO). International Classification of Disease, 11th Revision (ICD-11). WHO. 2018. URL: http://www.who.int/classifications/icd/en/ [accessed 2022-06-22]
49. Auriacombe M, Serre F, Denis C, Fatseas M. Diagnosis of addictions. In: The Routledge Handbook of the Philosophy and Science of Addiction. London, UK: Routledge; Jun 11, 2018:132-144.
50. Dong G, Potenza MN. A cognitive-behavioral model of Internet gaming disorder: theoretical underpinnings and clinical implications. J Psychiatr Res 2014 Nov;58:7-11 [FREE Full text] [doi: 10.1016/j.jpsychires.2014.07.005] [Medline: 25062755]
51. Wu L, Zhu L, Shi X, Zhou N, Wang R, Liu G, et al. Impaired regulation of both addiction-related and primary rewards in individuals with internet gaming disorder. Psychiatry Res 2020 Feb 22;286:112892. [doi: 10.1016/j.psychres.2020.112892] [Medline: 32114205]
52. Jo YS, Bhang SY, Choi JS, Lee HK, Lee SY, Kweon Y. Clinical characteristics of diagnosis for Internet Gaming Disorder: Comparison of DSM-5 IGD and ICD-11 GD diagnosis. J Clin Med 2019 Jun 28;8(7):945 [FREE Full text] [doi: 10.3390/jcm8070945] [Medline: 31261841]

53. Ko C, Yen J, Chen S, Wang P, Chen C, Yen C. Evaluation of the diagnostic criteria of Internet gaming disorder in the DSM-5 among young adults in Taiwan. J Psychiatr Res 2014 Jun;53:103-110. [doi: 10.1016/j.jpsychires.2014.02.008] [Medline: 24581573]

Abbreviations

2PL: 2-parameter logistic
CFA: confirmatory factor analysis
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
ICD: International Classification of Diseases
IGD: internet gaming disorder
IRT: Item Response Theory
ScUD: screen use disorder
WHO: World Health Organization

Edited by A Mavragani; submitted 06.07.21; peer-reviewed by M Haucke, L Maldonado-Murciano; comments to author 21.12.21; revised version received 26.02.22; accepted 08.06.22; published 27.07.22.

Please cite as:
Boudard M, Alexandre JM, Kervran C, Jakubiec L, Shmulewitz D, Hasin D, Fournet L, Rassis C, Claverie P, Serre F, Auriacombe M
Item Response Theory Analyses of Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) Criteria Adapted to Screen Use Disorder: Exploratory Survey
J Med Internet Res 2022;24(7):e31803
URL: https://www.jmir.org/2022/7/e31803
doi:10.2196/31803
PMID:35896018

Original Paper

A Google Trends Approach to Identify Distinct Diurnal and Day-of-Week Web-Based Search Patterns Related to Conjunctivitis and Other Common Eye Conditions: Infodemiology Study

Michael S Deiner[1,2], PhD; Gurbani Kaur[2,3], AB; Stephen D McLeod[1,2], MD; Julie M Schallhorn[1,2], MD, MS; James Chodosh[4,5], MPH, MD; Daniel H Hwang[6,7]; Thomas M Lietman[1,2,8,9], MD; Travis C Porco[1,2,8,9], MPH, PhD

[1]Francis I Proctor Foundation, University of California San Francisco, San Francisco, CA, United States
[2]Department of Ophthalmology, University of California San Francisco, San Francisco, CA, United States
[3]School of Medicine, University of California San Francisco, San Francisco, CA, United States
[4]Department of Ophthalmology, Harvard Medical School, Boston, MA, United States
[5]Massachusetts Eye and Ear, Harvard Medical School, Boston, MA, United States
[6]Stanford University, San Mateo, CA, United States
[7]The Nueva School, San Mateo, CA, United States
[8]Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA, United States
[9]Global Health Sciences, University of California San Francisco, San Francisco, CA, United States

Corresponding Author:
Travis C Porco, MPH, PhD
Department of Epidemiology and Biostatistics
University of California San Francisco
490 Illinois St, Floor 2, Box 0944
San Francisco, CA, 94158
United States
Phone: 1 415 476 4101
Email: travis.porco@ucsf.edu

Abstract

Background: Studies suggest diurnal patterns of occurrence of some eye conditions. Leveraging new information sources such as web-based search data to learn more about such patterns could improve the understanding of patients' eye-related conditions and well-being, better inform timing of clinical and remote eye care, and improve precision when targeting web-based public health campaigns toward underserved populations.

Objective: To investigate our hypothesis that the public is likely to consistently search about different ophthalmologic conditions at different hours of the day or days of week, we conducted an observational study using search data for terms related to ophthalmologic conditions such as conjunctivitis. We assessed whether search volumes reflected diurnal or day-of-week patterns and if those patterns were distinct from each other.

Methods: We designed a study to analyze and compare hourly search data for eye-related and control search terms, using time series regression models with trend and periodicity terms to remove outliers and then estimate diurnal effects. We planned a Google Trends setting, extracting data from 10 US states for the entire year of 2018. The exposure was internet search, and the participants were populations who searched through Google's search engine using our chosen study terms. Our main outcome measures included cyclical hourly and day-of-week web-based search patterns. For statistical analyses, we considered $P<.001$ to be statistically significant.

Results: Distinct diurnal ($P<.001$ for all search terms) and day-of-week search patterns for eye-related terms were observed but with differing peak time periods and cyclic strengths. Some diurnal patterns represented those reported from prior clinical studies. Of the eye-related terms, "pink eye" showed the largest diurnal amplitude-to-mean ratios. Stronger signal was restricted to and peaked in mornings, and amplitude was higher on weekdays. By contrast, "dry eyes" had a higher amplitude diurnal pattern on weekends, with stronger signal occurring over a broader evening-to-morning period and peaking in early morning.

Conclusions: The frequency of web-based searches for various eye conditions can show cyclic patterns according to time of the day or week. Further studies to understand the reasons for these variations may help supplement the current clinical understanding of ophthalmologic symptom presentation and improve the timeliness of patient messaging and care interventions.

(J Med Internet Res 2022;24(7):e27310) doi:10.2196/27310

KEYWORDS

diurnal eye conditions; hebdomadal; online search; web-based search; eye conditions; infodemiology; dry eye; conjunctivitis; pink eye; information seeking; vision

Introduction

Infodemiology is a relatively young discipline within health informatics studying the science of distribution and determinants of information within an electronic medium, specifically the internet or in a population, with the aim of informing public health and policy [1-3]. Applications of this form of health informatics have included predicting coronavirus outbreaks based upon queries of web-based search engines, syndromic surveillance by analysis of status updates or tweets on Twitter, tracking the disparities in access to health care information, and mining search engine data to cluster query click data to estimate prevalence of certain conditions that patients seek to address themselves outside of clinical settings or hours or to study prevalence of factors in typically unobserved locations [2,4-6]. One key advantage proffered by these approaches to public health analytics compared to collating and probing large data sets is the ability to conduct real time predictive analysis of health-related behaviors [2,7,8]. For example, one study found that the number of clicks on a keyword-triggered link in Google demonstrated a strong correlation with the following week of influenza cases during the 2004-2005 Canadian influenza season [9]. Similarly, another study found that social media–based surveillance for foodborne diseases were 66% as effective, rapid, and cheaper than standard database surveillance systems [10].

Google Trends has become a popular tool for infodemiologic studies in predicting disease occurrence and outbreaks, so much so that standardized approaches seeking to strengthen validity of such analyses have been proposed, and commonly used data access tools have been developed [11,12]. However, limitations when using Google Trends must also be considered. For example, for COVID-19, media coverage can affect web searches [13,14], and search volume values can vary depending on the date of data collection [15,16]. Applications in this field are vast (eg, use of Google Trends for public health planning regarding marginalized populations or birth control, to name a few [17,18]) and can adapt rapidly to current events [3]. Recent Google Trends studies have explored, for example, the potential impact of the COVID-19 pandemic on mental health behavior and child mistreatment [19-22] on ocular and other communicable and noncommunicable disease [23,24] and on treatment and misinformation related to COVID-19 itself [25-27].

Cyclic patterns of Google Trends search interest as related to human health, often seasonal but also to a letter extent diurnal, are an area of extensive research. Clinical study has identified cyclic occurrence of health conditions in humans, including diurnal eye-related conditions, and the results may facilitate chronopreventive and chronotherapeutic care [28-35]. Web-based search behavior regarding nonocular disease symptoms has been shown to reflect seasonal and diurnal clinical cyclicity as well as aspects of disease not typically observed in clinics at all (for example, coronary heart disease and depression) [36,37]. Web-based search or social media data also can reflect seasonal or emerging clinical eye disease patterns and conjunctivitis epidemics on relatively long timescales, including the impact of other factors such as the COVID-19 pandemic [8,23,38-43]. This suggested that, as with other health conditions [36,37], there is the potential to add to our knowledge about diurnal and day of week aspects of eye disease outside of the days and times that patients are typically seen in clinics, using web-based hourly search data. Herein, we tested the hypothesis that the public is likely to search about different aspects of eye health at different (but predictable) hours of the day or days of week. Specifically, we conducted an observational study investigating if US hourly web-based search data for terms related to conjunctivitis or other common eye conditions and treatments could demonstrate diurnal or day-of-week cyclic patterns and if those patterns were distinct from each other. For example, daily occurrence peaks may occur at different times, or the difference between the peak and the trough may differ.

Methods

Google Search Data

We queried Google Trends for conjunctivitis terms and other common eye conditions and treatments for comparison. Search terms included “conjunctivitis,” “blurry eyes,” “cataracts,” “pink eye,” “dry eyes,” “watery eyes,” “glaucoma,” “contact lenses,” “visine,” and “lasik.” A positive control term that would likely exhibit hourly and day-of-week variation (“drunk”) was included. Data were obtained using a Python (Python Software Foundation) script we developed to apply using *Pytrends* (a commonly used application programming interface to access Google Trends data) to obtain Historical Hourly Interest data, using the pytrends.get_historical_interest application programming interface [4,6,12]. Each term and state combination were queried individually. Each request retrieved 1 week of hourly data. The results were combined for analysis. Using this method, no categories were specified in the query, quotes were not used, and terms were queried individually as terms and not as *topics*. Relative search volume (RSV) of hourly search frequency data for these terms for the year 2018 (the most recent complete year of data available at the time of our query) from the 10 most populous US states (CA, FL, GA, IL, MI, NC, NY, OH, PA, and TX) were downloaded. Data were

queried and downloaded twice for each state-term pair to account for random sampling during the week of August 26, 2019 [36,37]. Universal Coordinated Times were adjusted to the predominant time zone for each state (only FL, MI, and TX include multiple time zones). The resulting time series represented hourly RSV for a given location, time period, and term. Data for all states were then combined for analysis.

Diurnal and Day of Week Analysis and Comparison of Cyclic Strength and Peak Times

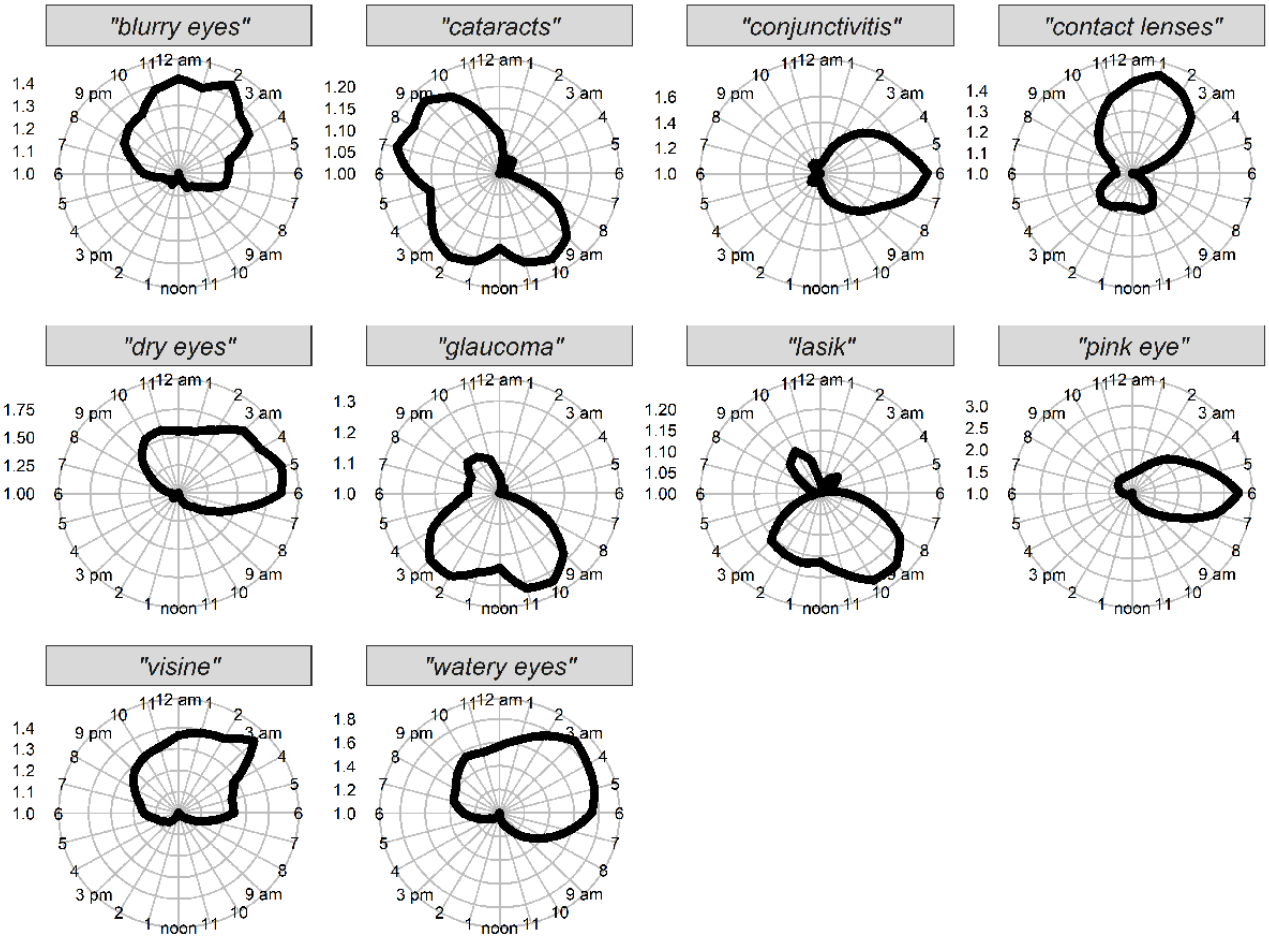
Using the hourly RSV for each search term as an outcome variable, we conducted Serfling regression adjusting for trend, as follows [44-46]. We adjusted for overall trend using third-order orthogonal polynomials in the number of days since January 1, 2018. Diurnal effects were modeled by terms of the form $\sin n\omega t$ and $\cos n\omega t$, where $\omega = 2\pi/24$, t is the time measured on a 24-hour clock, and n=1,…,4. We estimated separate diurnal effects for weekend days and for weekdays.

Because an outlier occurring at a single time could produce biased estimates of diurnal coefficients, we include additional terms to control for potential outliers (nuisance terms). We potentially include a large number of such terms of varying lengths, avoiding nonidentifiability by use of cross-validated LASSO (least absolute shrinkage and selection operator) to select only a small number of such terms [47]. This provides a simple regression-based filter for removing apparent epidemics and other irregular outliers. Specifically, outliers and localized (nonperiodic) departures were modeled by terms that take the value 1 on given intervals and are 0 otherwise. Specifically, we chose terms of the form $1x \in [km, k(m+1)-1)$, where m=0,1,…, x is the number of hours elapsed since midnight, January 1, 2018, and k takes values 8, 16, 32, 64, and 128, as well as 168 (the latter corresponding to the number of hours in a week). We also chose other intervals in a sensitivity analysis, finding that the choice of these regressors had little effect on the results; specifically, we chose the set k=7, 14, 28, 56, 112, and 168 hours, as well as the set 9, 18, 36, 72, 144, and 168 hours. Other choices for filtering outliers could have been chosen instead of this regression procedure.

For statistical analyses, following model selection for these nuisance terms, ordinary least squares estimation was used to estimate the trend, outlier, and trigonometric coefficients. From the trigonometric coefficients and intercept, we estimated the circular median occurrence time and the amplitude-to-mean ratio (in a similar fashion as our previous analyses and using the R [R Foundation for Statistical Computing] package "circular") [39-41]. Because diurnal and day of week occurrence data are angular data, we used the circular median time to summarize the central tendency; the circular median reflects the peak occurrence (when the data are approximately unimodal). The amplitude-to-mean ratio measures the cyclicity, with values near zero indicating small cyclic variability. Standard errors and P values were determined using time series bootstrap, with a fixed width of 20 hours [48-50]. For diurnal cyclic patterns, P values less than .001 were considered significant.

For data visualizations, mean hourly results of the filtered time series data for each term were normalized for visual comparison in polar plots (R package "ggplot" [51]). In order to optimally demonstrate cyclic patterns per terms in the plots, hourly RSVs for each term were normalized by dividing the mean per each hour per term by the value of the hour having the smallest mean value such that the hours with the least RSV are plotted closest to the center with a value of 1.0, while hours of higher relative search interest were plotted further from the center. Since values have been normalized, plots do not represent total search interest for one term vs another—but instead represent the relative amount of search interest between terms (Figure 1), between days for an individual term (Figure 2), or between seasons and weekday vs weekend day for an individual term (Figure 3).

Figure 1. Average hourly cyclic pattern from 2018 for 10 US states combined.



Ethics Approval

University of California San Francisco Institutional Review Board approval (14-14743) was obtained for this study.

Results

Overall, we found that each search term exhibited cyclic diurnal patterns of search interest (P<.001 for all terms). However, cyclic strength and central tendency differed between search terms, as described below.

Hourly, Weekly, and Seasonal Patterns

To visualize cyclic diurnal patterns for each term, mean RSV at each time of day is represented in normalized polar 24-hour plots (see Methods) in Figure 1. Note that despite most terms exhibiting diurnal patterns, scale bars in Figure 1 indicate that not all terms exhibited similar diurnal strength. In Figure 2, cyclic diurnal patterns for each term on each day of the week are presented as normalized polar 24-hour plots. These plots suggested some terms had diurnal cyclic features with patterns that varied between weekdays and weekends. Terms shown to have statistically significant day-of-week patterns, mean peak day values, and other day-of-week characteristics for all terms are shown in Table 1. In Figure 3, normalized polar 24-hour plots indicate cyclic diurnal patterns for each term for each season for weekday and weekends. Weekday group results are shown as solid lines, and weekend day group results as dashed lines. Seasons are indicated by color. These plots suggested that although most terms had similar diurnal and weekday search patterns per season, in some cases, features varied by season. For example, “dry eyes” tended to have more RSV overall throughout the hours of winter and spring, but also exhibited a strong morning peak seen in summer weekends, as did “watery eyes” in winter and spring weekends.”

Figure 2. Average hourly cyclic pattern per weekday from 2018 for 10 US states combined.

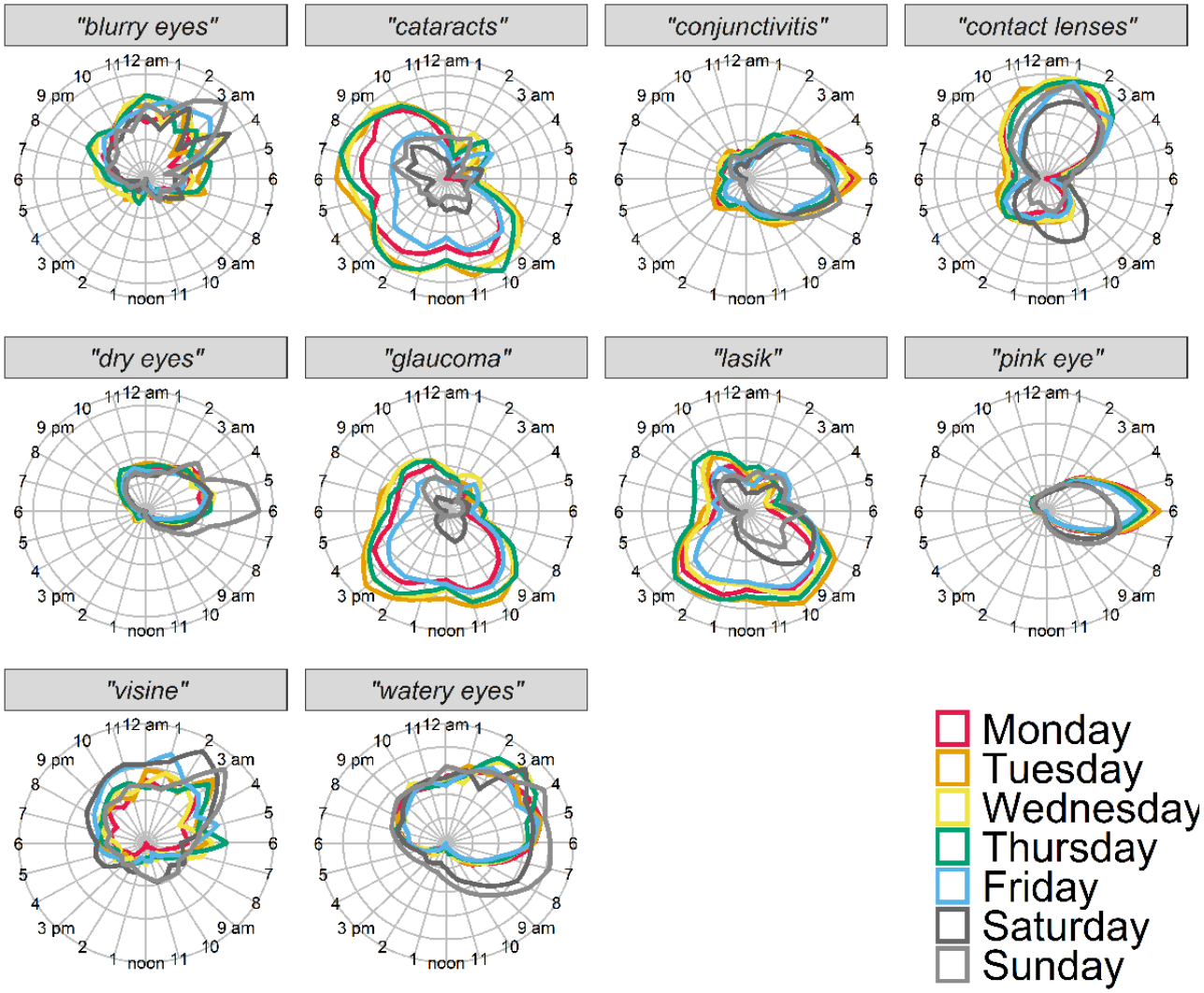


Table 1. Cyclical diurnal or day-of-week characteristics of relative search values.

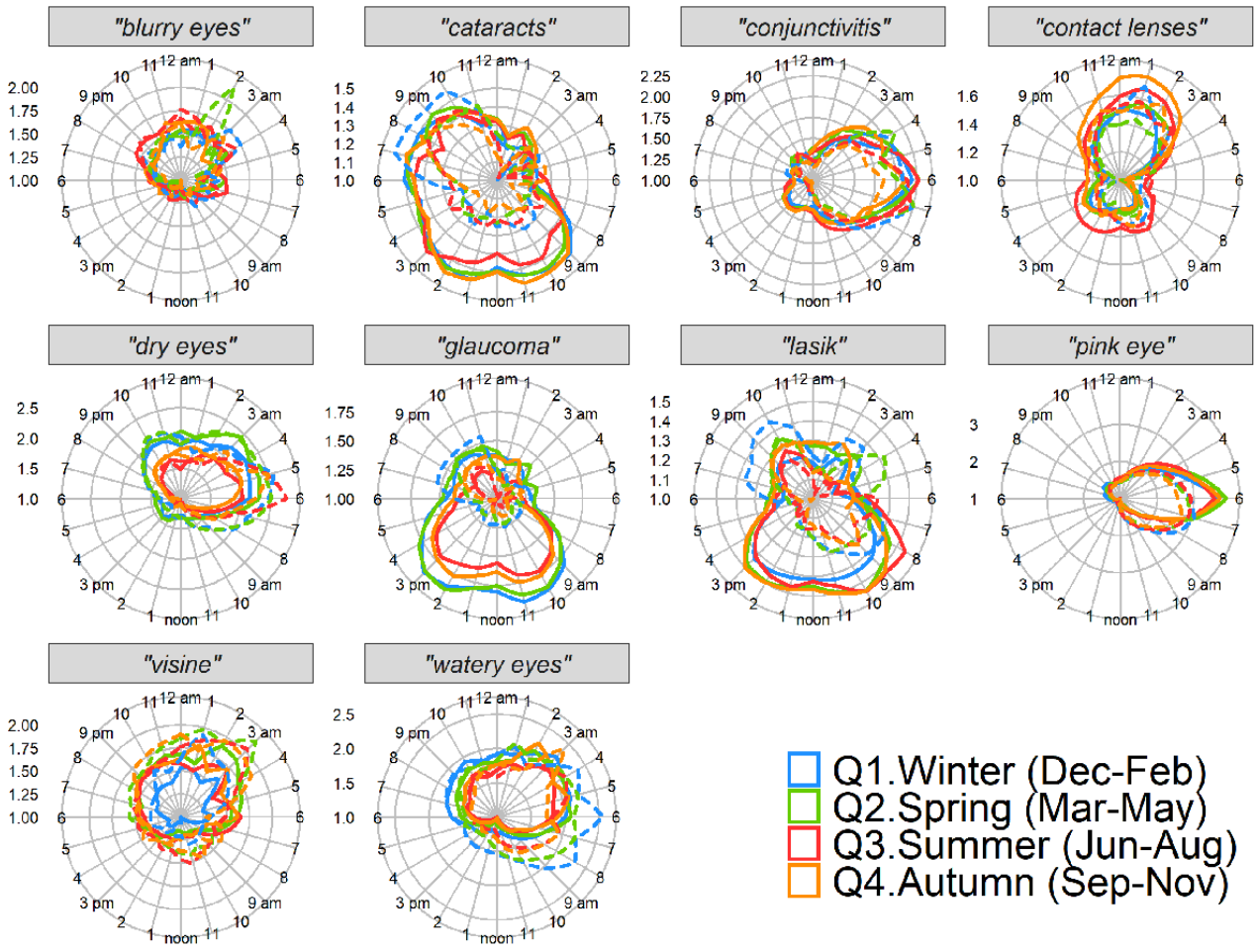
Search terms	Circular median times[a]			Amplitude-to-mean ratio		
	Weekday	Weekend	Difference, hrs	Weekday[b]	Weekend	Difference[c]
“blurry eyes”	00:39 (00:18, 00:59)	01:09 (00:40, 01:37)	0-2	0.37 (0.33, 0.47)	0.42 (0.39, 0.59)	0.02 (–0.04 to 0.08)
“cataracts”	15:28 (15:11, 15:45)	18:35 (17:12, 20:15)	2-4	0.26 (0.24, 0.3)	0.18 (0.15, 0.25)	–0.05 (–0.08 to –0.02)
“conjunctivitis”	06:11 (06:02, 06:21)	05:54 (05:46, 06:03)	0-2	0.62 (0.59, 0.66)	0.77 (0.72, 0.84)	0.04 (0 to 0.07)
“contact lenses”	23:16 (22:59, 23:32)	02:29 (01:57, 03:02)	2-4	0.43 (0.41, 0.47)	0.38 (0.35, 0.43)	–0.03 (–0.06 to –0.01)
“dry eyes”	02:13 (02:07, 02:20)	03:38 (03:28, 03:48)	0-2	0.61 (0.58, 0.66)	0.85 (0.79, 0.95)	0.12 (0.08 to 0.17)
“glaucoma”	13:24 (13:14, 13:34)	00:13 (21:41, 02:52)	4-6	0.38 (0.36, 0.41)	0.15 (0.12, 0.19)	–0.15 (–0.17 to –0.13)
“lasik”	11:51 (11:33, 12:09)	06:00 (05:10, 06:51)	4-6	0.28 (0.25, 0.31)	0.24 (0.2, 0.28)	–0.03 (–0.06 to –0.01)
“pink eye”	04:35 (04:33, 04:38)	05:38 (05:33, 05:43)	0-2	1.6 (1.56, 1.64)	1.08 (1.04, 1.11)	–0.26 (–0.29 to –0.24)
“visine”	00:37 (00:17, 00:56)	00:37 (00:08, 01:04)	0-2	0.41 (0.35, 0.5)	0.43 (0.38, 0.55)	0.03 (–0.03 to 0.09)
“watery eyes”	01:49 (01:41, 01:56)	03:50 (03:35, 04:07)	2-4	0.67 (0.63, 0.74)	0.48 (0.45, 0.58)	–0.07 (–0.12 to –0.03)

[a]The average filtered and detrended circular median time (and 95% CI) for each term for weekdays and weekend days. We found evidence that the CIs of the coefficients measuring diurnality excluded zero, indicating statistically significant diurnal variation (P<.001 for all values).

[b]Peak-to-trough divided by mean value to normalize the scalar difference. A larger average daily amplitude-to-mean ratios value indicates a more pronounced diurnal pattern.

[c]Allows a comparison of weekday to weekend diurnal cycle amplitude-to-mean ratios (ie, a comparison of cyclic strengths), providing the average difference (and 95% CI) between weekday vs weekend amplitude-to-mean ratios. Negative values indicate stronger weekday cyclic strength, and positive values indicate stronger weekend cyclic strength. Values further from 0 indicate a larger difference between weekdays and weekend days. This column is a difference in amplitudes divided by the average of the weekend and weekday means, not a difference between the previous two columns.

Figure 3. Average hourly cyclic pattern per season for weekdays (solid) and weekend days (dashed) from 2018 for 10 US states combined.



Statistical Analysis of Cyclic Patterns

Following smoothing and detrending, the resulting data set was used for all subsequent statistical analyses and data visualizations. The results for all terms are presented in Table 1. Columns 2-3 provide weekday and weekend circular median times, and column 4 provides differences between weekday and weekend for the peak times. Columns 5-6 provide amplitude-to-mean ratios. Column 7 provides a comparison of weekend to weekday cyclic ratios (values further from zero indicate larger differences, negative values indicate stronger weekday cyclic strength, and confidence intervals crossing zero indicate no significant difference).

For all terms, diurnal cyclic patterns were significant on weekdays and on weekends ($P<.001$ for all terms). Characteristics differed by search terms. Of the eye-related terms, "pink eye" had the strongest diurnal cyclic patterns based on amplitude-to-mean ratios, with stronger signal restricted to a narrow time window and peaking all mornings within the same 1-hour period. This showed a higher amplitude on weekdays (Table 1, columns 5-7; Figures 2 and 3). "Conjunctivitis" also had one of the stronger diurnal cyclic patterns, but lower than "pink eye," with a slightly later morning circular median time and less cyclic strength difference between weekend and weekday. In contrast to "pink eye," "dry eyes" exhibited a stronger diurnal pattern on weekends, with stronger signal occurring over a broader evening-to-morning time window, peaking in early morning and most significantly on Sunday mornings (Table 1, columns 5-7; Figure 2 and 3). Similar to "pink eye" though, "dry eyes" circular median times were nearby on weekday compared to weekend. By contrast, weekday vs weekend circular median times for "cataracts," "glaucoma," and "lasik" were less aligned, and weekday RSV was larger than weekend overall for these terms (Table 1, column 4; Figures 2 and 3). As a positive control, the term "drunk" exhibited a strong amplitude-to-mean ratio that was strongest from late evening through early morning on weekends (data not shown), reflecting late-evening alcohol consumption.

Discussion

Principal Findings

Web-based search behavior patterns for terms related to common eye conditions and treatments exhibited significant unique cyclic diurnal variation. This suggests that leveraging infodemiological approaches such as those demonstrated in this study can add information to our understanding of the times of day and night when different ocular conditions may be of the most or least perturbance or concern to patients. This may help augment our traditional understanding of ocular conditions, which has been based predominantly upon assessing patients with ophthalmology conditions during typical clinic hours. We observed that features occurring outside of typical clinic hours can differ between ophthalmologic condition–related search terms. For example, "pink eye" showed larger diurnal amplitude-to-mean ratios over a short daily weekday morning time period, while other terms such as "dry eyes" had a larger amplitude diurnal pattern on weekends, with stronger signal occurring over a broader evening-to-morning period compared to "pink eye."

For some individual search terms, we also found significant differences in diurnal search patterns on weekdays vs weekends for that term. This suggests infodemiological approaches can provide new understanding of specific days of the week, and hours of those days, on which particular ophthalmologic conditions are most affecting patients. Such approaches can add to the ongoing research studies to understand critical times or days for severity or treatment of symptoms and conditions outside of standard clinic hours for ocular conditions, as has also been studied for other disease [28,31-37,52-57]. We also observed that diurnal search patterns can differ by time of year for some terms, suggesting unique seasonal factors may affect the diurnal cycle of specific ocular conditions and raising the potential value of the approaches such as ours for enhancing the study of seasonal eye disease [29].

Limitations

Our study has potential for outliers, bias, and confounders. For example, an isolated event, such as a celebrity contracting conjunctivitis, could trigger an unusual search for "pink eye" at the time that the news story was reported. Similarly, an event such as a power outage could trigger temporary changes in search patterns. Furthermore, it is known that media coverage can impact search for COVID-19–related terms [13,14], and daily RSV can vary with data collection date [15,16]. To account for such aberrancies, we used a model with a regression-based filter to remove unusual surges or decreases such as isolated events and other irregular outliers (see Methods) and used averages from repeated queries and from multiple states and days to reduce potential imprecision. The results appeared stable in sensitivity analyses of the model. Although not incorrect, our approach used search keyword terms and did not allow query of search topics or health category; thus, the study of our health topics may not be fully complete. Future approaches comparing results from multiple years, and optimally using search topics and health categories for refinement in preliminary and final analyses, could provide additional model validation. Future applications of machine learning also has potential to improve the sensitivity and specificity of our model [11,58].

Despite these limitations, we found evidence in support of our model. For example, the results identified in our analyses often reflected components of known clinical understanding. The observed increase in hourly RSV from late night to early morning for "dry eyes" and "blurry eyes" is consistent with clinical reports of the symptoms [31,34]. Stronger amplitude for "dry eyes" observed on weekend mornings might represent elevated prior evening exposure to irritants such as smoke or alcohol, which have been reported to increase these symptoms [32,33]. The observed increase in hourly RSV from evenings to early mornings for "blurry eye," "contact lenses," "visine," and "watery eye" may reflect increased evening and nighttime symptoms of contact lens wearers [35]. Similarly, observed increases in hourly RSV in mornings for "conjunctivitis" and "pink eye" may reflect clinical findings as well [28]. In comparison, diurnal search for "cataracts," "glaucoma," and "lasik" occurred more during weekdays at daytimes. This

suggests information-seeking behavior related to ocular procedures or chronic conditions not associated with acute symptoms may be more likely to occur during the regular workday.

Conclusions

In this study, we establish evidence from web-based hourly search patterns that suggest there are distinct diurnal and weekly patterns undergirding web-based information-seeking behavior related to a variety of ophthalmologic symptoms and conditions. More precise temporal understanding of clinical eye disease presentation, hygiene, and health maintenance behaviors among patients outside of the clinic may be ascertained in the future through analysis of complementary data sources such as using web-based search data. This in turn could lead to improved approaches for diurnal eye disease monitoring and timing of resource allotment for ocular telemedicine, timely health care messaging, and clinical interventions.

Acknowledgments

This work was supported in part by grant 1R01EY024608-01A1 (Lietman, PI) from the National Institutes of Health National Eye Institute (NIH-NEI), grant EY002162 (Core Grant for Vision Research—Ullian, PI) from the NIH-NEI, grants EY013124 and EY021558 (Chodosh, PI) from the National Institutes of Health National Eye Institute (NIH-NEI), and an Unrestricted Grant from Research to Prevent Blindness (McLeod, PI). The sponsor or funding organization had no role in the design or conduct of this research.

Conflicts of Interest

None declared.

References

1. E-collection 'Infodemiology and Infoveillance'. JMIR Publications. URL: https://www.jmir.org/collection/view/69 [accessed 2021-08-10]
2. 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]
3. Mackey T, Baur C, Eysenbach G. Advancing Infodemiology in a Digital Intensive Era. JMIR Infodemiology 2022 Feb 14;2(1):e37115. [doi: 10.2196/37115]
4. Samaras L, Sicilia M, García-Barriocanal E. Predicting epidemics using search engine data: a comparative study on measles in the largest countries of Europe. BMC Public Health 2021 Jan 21;21(1):100 [FREE Full text] [doi: 10.1186/s12889-020-10106-8] [Medline: 33472589]
5. 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]
6. Hall J, Lo F, Saha S, Vaidyanathan A, Hess J. Internet searches offer insight into early-season pollen patterns in observation-free zones. Sci Rep 2020 Jul 09;10(1):11334 [FREE Full text] [doi: 10.1038/s41598-020-68095-y] [Medline: 32647115]
7. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):S154-S158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]
8. Tran HTT, Lu S, Tran HTT, Nguyen BV. Social Media Insights During the COVID-19 Pandemic: Infodemiology Study Using Big Data. JMIR Med Inform 2021 Jul 16;9(7):e27116 [FREE Full text] [doi: 10.2196/27116] [Medline: 34152994]
9. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. AMIA Annu Symp Proc 2006:244-248 [FREE Full text] [Medline: 17238340]
10. 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]
11. Mavragani A, Ochoa G. Google Trends in Infodemiology and Infoveillance: Methodology Framework. JMIR Public Health Surveill 2019 May 29;5(2):e13439 [FREE Full text] [doi: 10.2196/13439] [Medline: 31144671]
12. Pytrends: unofficial api for google trends. GitHub. URL: https://github.com/GeneralMills/pytrends/blob/master/README.md [accessed 2019-02-16]
13. Rovetta A. Reliability of Google Trends: Analysis of the Limits and Potential of Web Infoveillance During COVID-19 Pandemic and for Future Research. Front Res Metr Anal 2021 May 25;6:670226 [FREE Full text] [doi: 10.3389/frma.2021.670226] [Medline: 34113751]
14. Sousa-Pinto B, Anto A, Czarlewski W, Anto JM, Fonseca JA, Bousquet J. Assessment of the Impact of Media Coverage on COVID-19-Related Google Trends Data: Infodemiology Study. J Med Internet Res 2020 Aug 10;22(8):e19611 [FREE Full text] [doi: 10.2196/19611] [Medline: 32530816]

15. Sato K, Mano T, Iwata A, Toda T. Need of care in interpreting Google Trends-based COVID-19 infodemiological study results: potential risk of false-positivity. BMC Med Res Methodol 2021 Jul 18;21(1):147 [FREE Full text] [doi: 10.1186/s12874-021-01338-2] [Medline: 34275447]
16. Rabiolo A, Alladio E, Morales E, McNaught AI, Bandello F, Afifi AA, et al. Forecasting the COVID-19 Epidemic by Integrating Symptom Search Behavior Into Predictive Models: Infoveillance Study. J Med Internet Res 2021 Aug 11;23(8):e28876 [FREE Full text] [doi: 10.2196/28876] [Medline: 34156966]
17. Card KG, Lachowsky NJ, Hogg RS. Using Google Trends to Inform the Population Size Estimation and Spatial Distribution of Gay, Bisexual, and Other Men Who Have Sex With Men: Proof-of-concept Study. JMIR Public Health Surveill 2021 Nov 29;7(11):e27385 [FREE Full text] [doi: 10.2196/27385] [Medline: 34618679]
18. Prata N, Weidert K, Zepecki A, Yon E, Pleasants E, Sams-Abiodun P, et al. Using Application Programming Interfaces (APIs) to Access Google Data and Gain Insights Into Searches on Birth Control in Louisiana and Mississippi, 2014-2018: Infoveillance Study. J Med Internet Res 2021 Jul 12;23(7):e25923 [FREE Full text] [doi: 10.2196/25923] [Medline: 34255662]
19. Riem MME, De Carli P, Guo J, Bakermans-Kranenburg MJ, van IJzendoorn MH, Lodder P. Internet Searches for Terms Related to Child Maltreatment During COVID-19: Infodemiology Approach. JMIR Pediatr Parent 2021 Jul 13;4(3):e27974 [FREE Full text] [doi: 10.2196/27974] [Medline: 34174779]
20. Zitting K, Lammers-van der Holst HM, Yuan RK, Wang W, Quan SF, Duffy JF. Google Trends reveals increases in internet searches for insomnia during the 2019 coronavirus disease (COVID-19) global pandemic. J Clin Sleep Med 2021 Feb 01;17(2):177-184 [FREE Full text] [doi: 10.5664/jcsm.8810] [Medline: 32975191]
21. Wang A, McCarron R, Azzam D, Stehli A, Xiong G, DeMartini J. Utilizing Big Data From Google Trends to Map Population Depression in the United States: Exploratory Infodemiology Study. JMIR Ment Health 2022 Mar 31;9(3):e35253 [FREE Full text] [doi: 10.2196/35253] [Medline: 35357320]
22. 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 Jan 27;24(1):e32731 [FREE Full text] [doi: 10.2196/32731] [Medline: 34932494]
23. Deiner MS, Seitzman GD, Kaur G, McLeod SD, Chodosh J, Lietman TM, et al. Sustained Reductions in Online Search Interest for Communicable Eye and Other Conditions During the COVID-19 Pandemic: Infodemiology Study. JMIR Infodemiology 2022 Mar 16;2(1):e31732 [FREE Full text] [doi: 10.2196/31732] [Medline: 35320981]
24. Xu C, Cao Z, Yang H, Gao Y, Sun L, Hou Y, et al. Leveraging Internet Search Data to Improve the Prediction and Prevention of Noncommunicable Diseases: Retrospective Observational Study. J Med Internet Res 2020 Nov 12;22(11):e18998 [FREE Full text] [doi: 10.2196/18998] [Medline: 33180022]
25. Rovetta A. The Impact of COVID-19 on Conspiracy Hypotheses and Risk Perception in Italy: Infodemiological Survey Study Using Google Trends. JMIR Infodemiology 2021;1(1):e29929 [FREE Full text] [doi: 10.2196/29929] [Medline: 34447925]
26. Chejfec-Ciociano JM, Martínez-Herrera JP, Parra-Guerra AD, Chejfec R, Barbosa-Camacho FJ, Ibarrola-Peña JC, et al. Misinformation About and Interest in Chlorine Dioxide During the COVID-19 Pandemic in Mexico Identified Using Google Trends Data: Infodemiology Study. JMIR Infodemiology 2022 Jan 27;2(1):e29894 [FREE Full text] [doi: 10.2196/29894] [Medline: 35155994]
27. 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]
28. Smolensky MH, Portaluppi F, Manfredini R, Hermida RC, Tiseo R, Sackett-Lundeen LL, et al. Diurnal and twenty-four hour patterning of human diseases: acute and chronic common and uncommon medical conditions. Sleep Med Rev 2015 Jun;21:12-22. [doi: 10.1016/j.smrv.2014.06.005] [Medline: 25129839]
29. Qassim A, Viki M, Ng SK, Jersmann H, Casson RJ. Climate and season: the effects on ophthalmic diseases. Clin Exp Ophthalmol 2017 May;45(4):385-392. [doi: 10.1111/ceo.12883] [Medline: 27894161]
30. Sié A, Diarra A, Millogo O, Zongo A, Lebas E, Bärnighausen T, et al. Seasonal and Temporal Trends in Childhood Conjunctivitis in Burkina Faso. Am J Trop Med Hyg 2018 Jul;99(1):229-232 [FREE Full text] [doi: 10.4269/ajtmh.17-0642] [Medline: 29761759]
31. Guillon M, Shah S. Rationale for 24-hour management of dry eye disease: A review. Cont Lens Anterior Eye 2019 Apr;42(2):147-154. [doi: 10.1016/j.clae.2018.11.008] [Medline: 30497903]
32. Thomas J, Jacob GP, Abraham L, Noushad B. The effect of smoking on the ocular surface and the precorneal tear film. Australas Med J 2012;5(4):221-226 [FREE Full text] [doi: 10.4066/AMJ.2012.1035] [Medline: 22848314]
33. Kim JH, Kim JH, Nam WH, Yi K, Choi DG, Hyon JY, et al. Oral alcohol administration disturbs tear film and ocular surface. Ophthalmology 2012 May;119(5):965-971. [doi: 10.1016/j.ophtha.2011.11.015] [Medline: 22330961]
34. Walker PM, Lane KJ, Ousler GW, Abelson MB. Diurnal variation of visual function and the signs and symptoms of dry eye. Cornea 2010 Jun;29(6):607-612. [doi: 10.1097/ICO.0b013e3181c11e45] [Medline: 20458235]

35. Dumbleton KA, Guillon M, Theodoratos P, Patel T. Diurnal Variation in Comfort in Contact Lens and Non-contact Lens Wearers. Optom Vis Sci 2016 Aug;93(8):820-827. [doi: 10.1097/OPX.0000000000000920] [Medline: 27362615]
36. Tana JC, Kettunen J, Eirola E, Paakkonen H. Diurnal Variations of Depression-Related Health Information Seeking: Case Study in Finland Using Google Trends Data. JMIR Ment Health 2018 May 23;5(2):e43 [FREE Full text] [doi: 10.2196/mental.9152] [Medline: 29792291]
37. Senecal C, Widmer RJ, Lerman LO, Lerman A. Association of Search Engine Queries for Chest Pain With Coronary Heart Disease Epidemiology. JAMA Cardiol 2018 Dec 01;3(12):1218-1221 [FREE Full text] [doi: 10.1001/jamacardio.2018.3459] [Medline: 30422176]
38. Leffler CT, Davenport B, Chan D. Frequency and seasonal variation of ophthalmology-related internet searches. Can J Ophthalmol 2010 Jun;45(3):274-279. [doi: 10.3129/i10-022] [Medline: 20436544]
39. Deiner MS, Lietman TM, McLeod SD, Chodosh J, Porco TC. Surveillance Tools Emerging From Search Engines and Social Media Data for Determining Eye Disease Patterns. JAMA Ophthalmol 2016 Sep 01;134(9):1024-1030 [FREE Full text] [doi: 10.1001/jamaophthalmol.2016.2267] [Medline: 27416554]
40. Deiner MS, McLeod SD, Chodosh J, Oldenburg CE, Fathy CA, Lietman TM, et al. Clinical Age-Specific Seasonal Conjunctivitis Patterns and Their Online Detection in Twitter, Blog, Forum, and Comment Social Media Posts. Invest Ophthalmol Vis Sci 2018 Feb 01;59(2):910-920 [FREE Full text] [doi: 10.1167/iovs.17-22818] [Medline: 29450538]
41. Berlinberg EJ, Deiner MS, Porco TC, Acharya NR. Monitoring Interest in Herpes Zoster Vaccination: Analysis of Google Search Data. JMIR Public Health Surveill 2018 May 02;4(2):e10180 [FREE Full text] [doi: 10.2196/10180] [Medline: 29720364]
42. Deiner MS, McLeod SD, Wong J, Chodosh J, Lietman TM, Porco TC. Google Searches and Detection of Conjunctivitis Epidemics Worldwide. Ophthalmology 2019 Sep;126(9):1219-1229 [FREE Full text] [doi: 10.1016/j.ophtha.2019.04.008] [Medline: 30981915]
43. Deiner MS, Seitzman GD, McLeod SD, Chodosh J, Hwang DH, Lietman TM, et al. Ocular Signs of COVID-19 Suggested by Internet Search Term Patterns Worldwide. Ophthalmology 2021 Jan;128(1):167-169 [FREE Full text] [doi: 10.1016/j.ophtha.2020.06.026] [Medline: 32562704]
44. Serfling RE. Methods for current statistical analysis of excess pneumonia-influenza deaths. Public Health Rep 1963 Jun;78(6):494-506 [FREE Full text] [Medline: 19316455]
45. Serfling RE, Sherman IL, Houseworth WJ. Excess pneumonia-influenza mortality by age and sex in three major influenza A2 epidemics, United States, 1957-58, 1960 and 1963. Am J Epidemiol 1967 Sep;86(2):433-441. [doi: 10.1093/oxfordjournals.aje.a120753] [Medline: 6058395]
46. Wang X, Wu S, MacIntyre CR, Zhang H, Shi W, Peng X, et al. Using an adjusted Serfling regression model to improve the early warning at the arrival of peak timing of influenza in Beijing. PLoS One 2015 Mar 10;10(3):e0119923 [FREE Full text] [doi: 10.1371/journal.pone.0119923] [Medline: 25756205]
47. Heinze G, Wallisch C, Dunkler D. Variable selection - A review and recommendations for the practicing statistician. Biom J 2018 May 02;60(3):431-449 [FREE Full text] [doi: 10.1002/bimj.201700067] [Medline: 29292533]
48. Kreiss J, Lahiri S. Bootstrap methods for time series. In: Time Series Analysis: Methods and Applications. Oxford, UK: El-Sevier; 2012:3-26.
49. Vanalli C, Casagrandi R, Gatto M, Bevacqua D. Shifts in the thermal niche of fruit trees under climate change: The case of peach cultivation in France. Agricultural and Forest Meteorology 2021;300:108327 [FREE Full text] [doi: 10.1016/j.agrformet.2021.108327]
50. Deiner MS, Fathy C, Kim J, Niemeyer K, Ramirez D, Ackley SF, et al. Facebook and Twitter vaccine sentiment in response to measles outbreaks. Health Informatics J 2019 Sep 17;25(3):1116-1132 [FREE Full text] [doi: 10.1177/1460458217740723] [Medline: 29148313]
51. Wickham H. Ggplot2: elegant graphics for data analysis. Houston, Texas, USA: Springer; 2016.
52. de Jong B, Brandi-Dohrn F, van der Meulen IJE, van den Berg TJTP, Wacker K. Diurnal Variation in Straylight in Patients With Fuchs Endothelial Corneal Dystrophy and Controls. Cornea 2022 Feb 05. [doi: 10.1097/ICO.0000000000003002] [Medline: 35120348]
53. Yang Y, Zhang X, Chen Z, Wei Y, Ye Q, Fan Y, LiGHT China Trial Study Group. Intraocular pressure and diurnal fluctuation of open-angle glaucoma and ocular hypertension: a baseline report from the LiGHT China trial cohort. Br J Ophthalmol 2022 Jan 27:bjophthalmol-2021-320128 [FREE Full text] [doi: 10.1136/bjophthalmol-2021-320128] [Medline: 35086806]
54. Alenezi H, Ozkan J, Willcox M, Parnell G, Carnt N. Differential gene expression of the healthy conjunctiva during the day. Cont Lens Anterior Eye 2021 Jul 24:101494. [doi: 10.1016/j.clae.2021.101494] [Medline: 34315655]
55. Zander DB, Böhringer D, Fritz M, Grewing V, Maier PC, Lapp T, et al. Hyperosmolar Eye Drops for Diurnal Corneal Edema in Fuchs' Endothelial Dystrophy: A Double-Masked, Randomized Controlled Trial. Ophthalmology 2021 Nov;128(11):1527-1533 [FREE Full text] [doi: 10.1016/j.ophtha.2021.04.015] [Medline: 33892048]
56. McGlumphy EJ, Mihailovic A, Ramulu PY, Johnson TV. Home Self-tonometry Trials Compared with Clinic Tonometry in Patients with Glaucoma. Ophthalmol Glaucoma 2021 Nov;4(6):569-580. [doi: 10.1016/j.ogla.2021.03.017] [Medline: 33845191]

57. Chakraborty R, Micic G, Thorley L, Nissen T, Lovato N, Collins M, et al. Myopia, or near-sightedness, is associated with delayed melatonin circadian timing and lower melatonin output in young adult humans. Sleep 2021 Mar 12;44(3):zsaa208. [doi: 10.1093/sleep/zsaa208] [Medline: 33030546]
58. E-collection 'Infoveillance, Infodemiology, Digital Disease Surveillance, Infodemic Management'. JMIR Publications. URL: https://publichealth.jmir.org/collection/view/304 [accessed 2021-08-10]

Abbreviations

LASSO: least absolute shrinkage and selection operator
RSV: relative search volume

Edited by G Eysenbach; submitted 22.01.21; peer-reviewed by A Rovetta, A Mavragani; comments to author 26.03.21; revised version received 18.08.21; accepted 06.05.22; published 05.07.22.

Please cite as:
Deiner MS, Kaur G, McLeod SD, Schallhorn JM, Chodosh J, Hwang DH, Lietman TM, Porco TC
A Google Trends Approach to Identify Distinct Diurnal and Day-of-Week Web-Based Search Patterns Related to Conjunctivitis and Other Common Eye Conditions: Infodemiology Study
J Med Internet Res 2022;24(7):e27310
URL: https://www.jmir.org/2022/7/e27310
doi:10.2196/27310
PMID:35537041

Original Paper

Perceptions of Oral Nicotine Pouches on Reddit: Observational Study

Yihan Shao[1*], BA, BSc; Jonathan Zou[1*]; Zidian Xie[2], PhD; Rachel Grana Mayne[3], PhD; Deborah J Ossip[4], PhD; Irfan Rahman[5], PhD; Scott McIntosh[4], PhD; Dongmei Li[2], PhD

[1]Goergen Institute for Data Science, University of Rochester, Rochester, NY, United States

[2]Department of Clinical & Translational Research, University of Rochester Medical Center, Rochester, NY, United States

[3]Tobacco Control Research Branch, National Cancer Institute, Bethesda, MD, United States

[4]Department of Public Health Sciences, University of Rochester Medical Center, Rochester, NY, United States

[5]Department of Environmental Medicine, University of Rochester Medical Center, Rochester, NY, United States

[*]these authors contributed equally

Corresponding Author:
Dongmei Li, PhD
Department of Clinical & Translational Research
University of Rochester Medical Center
CU 420708
265 Crittenden Boulevard
Rochester, NY, 14642
United States
Phone: 1 5852767285
Email: Dongmei_Li@urmc.rochester.edu

Abstract

Background: Oral nicotine pouches are a new form of tobacco-free nicotine products launched in recent years with a variety of flavors.

Objective: This study aims to examine the public perceptions and discussions of oral nicotine pouches on Reddit, a popular social media platform for sharing user experiences.

Methods: Between February 15, 2019, and February 12, 2021, a total of 2410 Reddit posts related to oral nicotine pouches were obtained over a 2-year period. After the removal of unrelated or commercial posts, 653 Reddit posts related to oral nicotine pouches remained. Topics and sentiments related to oral nicotine pouches on Reddit were hand coded.

Results: The number of Reddit posts related to oral nicotine pouches increased during the study period. Content analysis showed that the most popular topic was “sharing product information and user experience” (366/653, 56%), in which sharing oral nicotine pouch products and user experiences were dominant. The next popular topic was “asking product-related questions” (product properties and product recommendations; 115/653, 17.6%), followed by “quitting nicotine products” such as vaping or smoking through use of oral nicotine pouches or quitting the oral nicotine pouches themselves (83/653, 12.7%) and “discussing oral nicotine pouch–related health” symptoms or concerns related to oral nicotine pouches (74/653, 11.3%). The least popular topic was “legality and permissions” related to oral nicotine pouches (15/653, 2.3%). In addition, a greater number of Reddit posts described positive attitudes compared to negative attitudes toward oral nicotine pouches (354/653, 54.2% vs 101/653, 15.5%; $P<.001$).

Conclusions: Reddit posts overall had a positive attitude toward oral nicotine pouches and users were actively sharing product and user experiences. Our study provides the first insight on up-to-date oral nicotine pouch discussions on social media.

(J Med Internet Res 2022;24(7):e37071) doi:10.2196/37071

KEYWORDS

oral nicotine pouches; Reddit; perception; nicotine; social media; sentiment; public opinion; user experience; attitude; content analysis; tobacco; smoking; cessation; quit; smoker; information seeking; information sharing; vaping

Introduction

Oral nicotine pouches are popular products launched by tobacco manufacturers over the past 3 years [1]. In appearance, oral nicotine pouches are similar to snus, a smokeless tobacco product originated in Sweden. Instead of containing the tobacco leaf found in snus, they are composed of white nicotine-containing powder [2]. The properties of being spit-free and tobacco-free may make oral nicotine pouches attractive alternatives to smokeless tobacco products. A cross-sectional survey study found that almost all oral nicotine pouch users were either former or current tobacco users (96.3%) [3]. The market for oral nicotine pouches is rapidly growing, indicated by the sales volume of the global market, which is estimated at US $2.33 billion in 2020 and expected to reach US $21.84 billion by the end of 2027 [4]. Among a great variety of brands in the United States, ZYN, a Swedish Match product, by far occupied the largest national market share (nearly two-thirds), and shipments of ZYN in the United States have increased from 114.1 million cans in 2020 to 173.9 million cans in 2021 [5].

To date, several studies have investigated the chemical components of oral nicotine pouches. For example, a recent study examined the moisture content, pH, and nicotine content of 37 oral nicotine pouch brands [6]. They found a similar pH across products and nicotine content comparable to, or greater than, smokeless tobacco pouch products, suggesting a similar addiction potential as those products. Due to the high nicotine content, the authors expressed concern over appeal to, and use by, young people and that oral nicotine pouches may be a gateway to more harmful tobacco products. The various flavors contained in the oral nicotine pouches are another rising concern as flavors are one of the major reasons that youth are attracted to and use electronic nicotine delivery system products with high nicotine content [7-9]. Nevertheless, another study estimated the toxicant levels of oral nicotine pouches and concluded that users are likely to be exposed to lower levels of toxic compounds than with Swedish snus [10]. Given that Swedish snus has been shown to have significantly fewer health risks than cigarette smoking, these results indicate that nicotine pouch users may be at lower risk of adverse health effects than users of Swedish snus or cigarettes [10]. Potential health effects of oral nicotine pouches await further investigation. Although there have been studies about the chemical characteristics of oral nicotine pouches [6,10,11], few studies have investigated public perceptions of these products, the importance of which is highlighted further by the substantial and increasing use of oral nicotine pouches.

The universality of social media makes it a suitable platform to collect information on public attitudes and perceptions of new emerging trends. Reddit is a popular social media outlet that is differentiated from other social media platforms by its capability for users to form niche groups called “subreddits.” Each subreddit has a common topic and users within the subreddit can express their opinions and personal experiences. According to *The Wall Street Journal*, there are over 52 million daily active users of Reddit [12]. The posts on Reddit tend to have higher character limits, allowing users to provide more information on opinions and experiences, as well as have meaningful discussions with other users that have similar interests. Thus, Reddit posts on oral nicotine pouches contain more information on user opinions and product use experiences than posts on other social media platforms. A number of previous studies have used Reddit to examine user perceptions of other tobacco products, for example, public perceptions of different e-cigarettes on Reddit [13,14].

In this study, we aimed to examine the public perceptions of, and popular topics related to, oral nicotine pouches on Reddit during the past 2 years, when the sales of oral nicotine pouches significantly increased in the United States. First, to understand what is on market, we identified the brands and flavors of oral nicotine pouch products available online. Through manually coding Reddit posts for sentiments and topics, we examined public attitudes toward oral nicotine pouches and identified top topics related to oral nicotine pouches. Classifications of subtopics and analysis of sentiments within each topic were further conducted to obtain more details. The findings provide important information on this emerging nicotine product, which could be helpful for understanding perception, use, and future regulation of oral nicotine pouches.

Methods

Online Collection of Oral Nicotine Pouches

Coders (YS and JZ) searched online stores in February 2021 to identify oral nicotine pouch products and brands on the market during the Reddit data collection period and to inform the hand coding of the Reddit posts. Information about available oral nicotine pouch products, including brand names, product names, and flavors, was collected from online stores like “nicokik.com” and “snusdirect.com.” The nicotine strengths, when available, were also included.

Data Collection and Preprocessing

Reddit data between February 15, 2019, and February 12, 2021, were downloaded from Reddit Archive [15] in March 2021. By keyword searching, 2410 posts containing keywords related to oral nicotine pouches and brand names, including “nicotine pouch,” “oral nicotine pouch,” “loop nicotine,” “lyft nicotine,” “ordic spirit pouch,” “zyn,” “velo,” “killapods,” “zonex,” “2one,” and “on!” were further extracted. Information including author, subreddit, created time, URL, and texts of the posts were included in our final data set.

To achieve higher accuracy of hand coding and defining the topics, we adopted an inductive approach [16]. We considered the method of content analysis, which is a research tool used to identify the presence of certain words, themes, or ideas in qualitative data [17]. We first randomly selected a sample of 350 posts (15% of the 2410 total posts), which was used to develop our initial codebook. Based on whether content was primarily focused on oral nicotine pouches or not, posts were first classified into related posts or unrelated posts. Unrelated posts were Reddit posts that did not mention oral nicotine pouches. Then, among related posts, we further classified them into commercial or noncommercial posts. Commercial posts were those directly posted by oral nicotine pouch sellers on Reddit, which contained advertisements for a specific product

in text or had the URL directly linked to online stores of oral nicotine pouches. We focused our analysis on the related and noncommercial posts. Among the 350 posts, 233 of them were unrelated or commercial, and 117 posts were related and noncommercial. By applying the same rules to all posts, in total, 27.1% (653/2410) posts were related and noncommercial, 2.3% (56/2410) posts were commercial, and 70.6% (1701/2410) posts were unrelated (Multimedia Appendix 1).

Topic Analysis

By examining the content of the 117 related and noncommercial posts, two authors (YS and JZ) individually grouped those with similar content and summarized them by a generalizable phrase such as "Sharing product information and user experience." Within each general topic, we further summarized analogous posts by a more specific topic (subtopic) like "user experiences." After hand coding 117 sampled posts separately, we compared and discussed our codebook definition from two individual coders, and the final decision on how to categorize topics was made by a group of 4 authors (YS, JZ, ZX, and DL) to form a unified classification standard. The initial codebook was used as the reference for the remaining 2060 posts, which were independently coded by 2 authors (YS and JZ). Any coding difference between the 2 coders was resolved through discussion among all 4 coders. Moreover, further modifications to the codebook were made in the process of hand coding the remaining 2060 posts with intensive discussions among the group of 4 coders. As a result, Reddit posts related to oral nicotine pouches were categorized into 5 topics: "sharing product information and user experience," "asking oral nicotine pouch–related questions," "quitting nicotine products," "discussing oral nicotine pouch–related health," and "legality/permissions" (Table 1). Each post was only assigned one topic or subtopic.

Table 1. Topics and subtopics of Reddit posts on oral nicotine pouches.

Topics and subtopics	Description
Sharing product information and user experience	
Products	Posts where users were sharing about specific oral nicotine pouch products.
User experiences	Posts where users were sharing the experience of how they were using oral nicotine pouches and how they felt about them when they were using the products.
Opinions	Posts where users were sharing subjective points of view about the products.
Information	Posts where users were sharing information that they obtained elsewhere.
Others	Posts about other things like sharing homemade oral nicotine product recipes.
Asking oral nicotine pouch–related questions	
Characteristics of oral nicotine pouches	Posts about the properties of oral nicotine pouches such as nicotine strength, flavors, and packaging.
Product recommendation	Posts that were asking for recommendations of oral nicotine pouch products.
Order or purchase information	Posts asking for information about the online and offline purchase of oral nicotine pouches.
Others	Posts asking uncommon questions, like asking for opinions about the future of the oral nicotine pouch market.
Quitting nicotine products	
Using oral nicotine pouches to quit vaping	Posts by users who were discussing intentions to use, attempts to use, or successful use of certain oral nicotine pouch products to quit vaping
Using oral nicotine pouches to quit smoking	Posts by users who were discussing intentions to use, attempts to use, or successful use of certain oral nicotine pouch products to quit smoking
Quitting oral nicotine pouches	Users of the posts intended to, were attempting to, or have successfully quit oral nicotine pouches
Others	Posts that shared experience using oral nicotine pouches to quit other nicotine-related products like tobacco and nicotine itself
Discussing oral nicotine pouch–related health	
Health symptoms	Posts discussing having negative health symptoms after the consumption of oral nicotine pouches, like gum bleeding and sickness
Health concerns	Posts that expressed health concerns with oral nicotine pouches. Users of those posts expressed worries about the potential side effects of oral nicotine pouches, including the symptoms above and addiction to nicotine.
Legality and permissions	Posts that were discussing the regulatory policies and legal issues related to oral nicotine pouches

Sentiment Analysis

In addition to analyzing the major topics discussed in each post, we performed sentiment analysis with the standard "positive," "neutral," and "negative" categories [18-22]. Example Reddit posts in each topic and subtopic category with different sentiments are shown in Multimedia Appendix 2. If the post expressed an overall positive attitude toward one of the oral

nicotine pouch products or the oral nicotine pouches as a whole, we considered it as positive. For example, posts where users shared enjoyable user experiences with one of the oral nicotine pouch products or found oral nicotine pouches helpful for quitting other tobacco products were considered positive. On the other hand, if the post expressed an overall negative attitude, we considered it negative. For example, posts discussing having adverse health effects after using oral nicotine pouches or being pessimistic about the products' future appeal were categorized as negative. Finally, if the post expressed neither an overall positive nor negative attitude, we labeled it neutral.

All the posts were double-coded. By focusing on 653 related and noncommercial posts, we calculated the agreement rates between topics and sentiments to examine the overall accuracy of classifications. The raw agreement rate (number of agreed posts/total posts) is 86.23% (563/653) for topics and 82.20% (537/653) for sentiments. The Cohen κ is 0.7526 for topics and 0.7102 for sentiments.

Ethics Approval

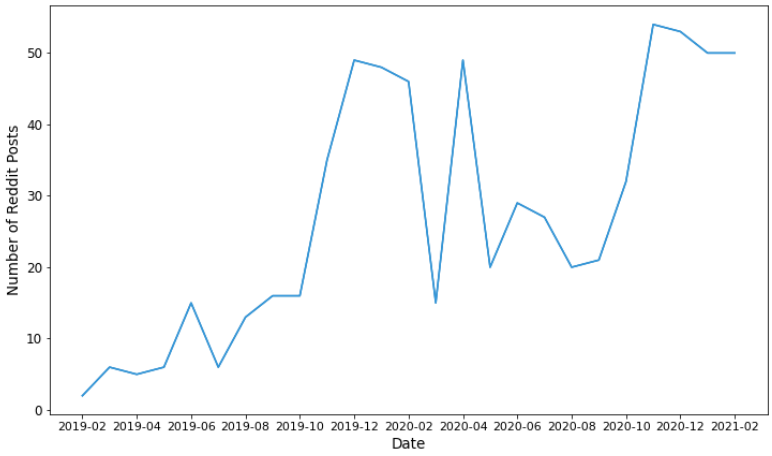
The study has been reviewed and approved by the Office for Human Subject Protection Research Subjects Review Board (RSRB) at the University of Rochester (Study ID: STUDY00006570).

Results

Longitudinal Trend of Number of Reddit Posts

In total, coders (YS and JZ) identified 20 oral nicotine pouch brands and 30 different oral nicotine products available on online stores during the study period (Multimedia Appendix 3). To examine the trend in discussions about oral nicotine pouches on Reddit, we plotted the total number of posts per month from February 2019 to February 2021. As shown in Figure 1, there has been a clear upward trend in the number of posts related to oral nicotine pouches, with the most significant rise occurring from February 2019 (2 posts) to December 2019 (49 posts) and August 2020 (20 posts) to November 2020 (54 posts). Although the increase was steady in the two time periods mentioned above, very large fluctuations appeared between December 2019 and August 2020. Within that period, the number of posts peaked in April 2020.

Figure 1. Longitudinal trend of Reddit posts related to oral nicotine pouches.



Topics and Subtopics

To understand the major topics discussed related to oral nicotine pouches on Reddit, we first divided posts into 5 general categories. Over half of the posts (366/653, 56%) were related to the topic "sharing product information and user experience." The second most popular topic (115/653, 17.6%) was "asking oral nicotine pouch–related questions," followed by "quitting nicotine products" (83/653, 12.7%) and "discussing oral nicotine pouch–related health" (74/653, 11.3%). Lastly, 2.3% (15/653) of posts reviewed the topic "legality and permissions."

Among the 5 topics, 4 were further divided into subtopics. Under the topic "sharing product information and user experience," users were more inclined to share "products" (140/366, 38.3%). Sharing "user experiences" accounted for the second-largest proportion (109/366, 29.8%), followed by sharing "opinions" (60/366, 18.6%). Sharing "information" and sharing "others" accounted for 10.7% (39/366) and 2.7% (10/366), respectively. For the topic "asking oral nicotine pouch–related questions," almost half of the posts (57/115, 49.6%) were asking about "characteristics" of oral nicotine pouches. Approximately one-quarter (27/115, 23.5%) were asking for "recommendations." Asking about "others" and asking for "order or purchase information" comprised 17.4% (20/115) and 9.6% (11/115), respectively. For the topic "quitting nicotine products," over half of the discussions (42/83, 50.6%) were about "using oral nicotine pouches to quit vaping." The remaining 3 subtopics were "using oral nicotine pouches to quit smoking" (15/83, 18.1%), quitting "others" (14/83, 16.9%), and "quitting oral nicotine pouches" (12/83, 14.5%). The fourth significant topic was "discussing oral nicotine pouch–related health." When discussing health, the majority of posts (60/74, 81.1%) were talking about "health symptoms" related to oral nicotine pouches, and the rest talked about either "health concerns" (13/74, 17.6%) about oral nicotine pouches or "others" (1/74, 1.4%).

Sentiments

Sentiment analysis of the Reddit posts determined that over half of the posts (354/653, 54.2%) expressed positive attitudes

toward oral nicotine pouches (Table 2). Approximately 30% of the posts (198/653, 30.3%) showed neutral attitudes, while 15.5% (101/653) conveyed negative attitudes toward oral nicotine pouches. The proportion of positive posts was significantly higher than that of negative posts (*P* value of the two-proportion *z* test was <.001). With respect to each topic, "quitting" contained the most significant proportion of positive posts (57/83, 68.7%). "Product information/user experience" was also dominated by positive posts (233/366, 63.7%). On the other hand, the topic "health" had the highest proportion of negative posts (54/74, 73%) within all 5 topics. "Legality and permissions" had only 13.3% (2/15) negative posts. In the case of neutral posts, the topics "product-related questions" and "legality and permissions" contained the highest proportion (62/115, 53.9% and 8/15, 53.3%) of posts.

Table 2. Sentiments toward oral nicotine pouches in Reddit posts.

Topic	Positive, n/N (%)	Neutral, n/N (%)	Negative, n/N (%)
Sharing product information and user experience	233/366 (63.7)	99/366 (27)	34/366 (9.3)
Product-related questions	48/115 (41.7)	62/115 (53.9)	5/115 (4.4)
Quitting nicotine products	57/83 (68.7)	20/83 (24.1)	6/83 (7.2)
Discussing oral nicotine pouch–related health	11/74 (14.9)	9/74 (12.1)	54/74 (73)
Legality and permissions	5/15 (33.3)	8/15 (53.3)	2/15 (13.4)

Discussion

Principal Findings

This study explored user perceptions of oral nicotine pouches on Reddit by performing topic and sentiment analysis on 653 Reddit posts related to oral nicotine pouches collected over 2 years. We observed an overall increasing trend in the number of posts during the study period. Major topics about oral nicotine pouches include sharing preferred products, user experiences, or opinions of oral nicotine pouches; quitting oral nicotine pouches or using oral nicotine pouches to quit other tobacco products; and adverse health effects. Over half of the posts showed a positive attitude toward oral nicotine pouches. Although the majority of product information/user experience and quitting-related posts indicated positive sentiments similar to the overall sentiment, health-related posts contained the most significant proportion of negative posts.

Comparison With Prior Work

It is interesting to note that the overall increase in the number of Reddit posts is consistent with the remarkable growth trend of sales of oral nicotine pouches during a similar time period. Compared to the sales of total other tobacco products in convenience stores in the United States for the 24 weeks ending on May 30 in 2020, which had a growth of 7%, sales of oral nicotine pouches grew a staggering 498% [23]. With regard to the striking fluctuations between February 2020 and August 2020, a potential and critical factor could be the global COVID-19 pandemic. The outbreak of COVID-19 constrained the tobacco product market in 2020 due to supply chains being disrupted by trade restrictions [24]. In addition, consumption and demand also declined because of lockdowns imposed by governments. It is therefore likely that the negative impact of the COVID-19 pandemic on the market and the public's distraction by the pandemic contributed to the fluctuating interest in oral nicotine pouches reflected in the sudden rise and fall of the number of posts on the public subreddits included in our study.

Several factors could be responsible for the large proportion of posts discussing topics related to product information/user experience information of oral nicotine pouches. One factor could be that oral nicotine pouches are novel products, generating curiosity. Another factor could be the large variety of oral nicotine pouches brands and flavors, which further engenders curiosity, interest, and willingness to share their preferences and user experiences with others. Moreover, compared to other social media platforms like Instagram or Twitter, Reddit has a particular feature called a "subreddit," where all users in the same subreddit are interested in the same topic and are only allowed to discuss related subjects [25]. For example, posts related to oral nicotine pouches were largely collected under subreddit like "NicotinePouches" and "zyn." These oral nicotine pouch–related subreddits facilitated the discussions on oral nicotine pouches, which helped us understand the attitude and topics related to oral nicotine pouches through these commonly used platforms. However, due to the recent emergence of oral nicotine pouches, the subreddits "Snus" and "Nicotine" also contained relevant posts. This topic identity would create a more familiar atmosphere for users with the same interest, thus facilitating more active and positive sharing.

In this study, the overall sentiment toward oral nicotine pouches was positive, and most quitting-related posts were positive as well, which is consistent with Andersson and Lundqvist's [1] finding that attitudes were more positive toward oral nicotine pouches than tobacco-containing snus [26]. The prevalence of positive posts relating to quitting smoking and vaping by switching to oral nicotine pouches might be due to misperceptions of the oral nicotine pouch products. Content analysis of quitting-related posts indicated that users (who had concerns about harmful effects of tobacco products such as cigarettes or e-cigarettes) considered the oral nicotine pouches as a "safer option" to help them quit vaping, as oral nicotine pouches contain sufficient nicotine content and provide a similar "buzz" as cigarettes or e-cigarettes. Some users may perceive the oral nicotine pouch products as useful for smoking cessation as a less hazardous form of nicotine, similar to nicotine replacement medication. Additionally, Struik and Yang [27]

conducted a content analysis of a quit vaping community on Reddit and discovered that among 175 posts that mentioned methods of quitting, 13 users reported using nicotine pouches to help them quit e-cigarettes. In addition, those perceptions may be influenced by recent events in the oral nicotine product marketplace. In 2009, Reynolds American Inc (then the second-largest tobacco company in the United States) acquired Niconovum AB and in 2012 began test-marketing their nicotine replacement therapy products in the United States (Zonnic pouches and Zonnic mini lozenges) after obtaining approval from the US Food and Drug Administration (FDA) Center for Drug Research and Evaluation. However, in 2019, despite a national release and marketing in major convenience store chains, the company pulled the products from the US market for undisclosed reasons [28]. Currently, British American Tobacco, which owns Reynolds American and its subsidiaries [29,30], markets oral nicotine products (Velo, a nicotine pouch, and Revel, a nicotine lozenge) as commercial tobacco products [31], and has submitted a Premarket Tobacco Product Application to the FDA Center for Tobacco Products for the Velo product [32]. The proliferation of these products in the marketplace and the former availability of the Zonnic product as an FDA-approved cessation aid might engender confusion that using oral nicotine pouches can help with tobacco cessation [33,34]. In addition, the marketing of oral nicotine pouches as "tobacco-free" might convey to users that oral nicotine pouches are low-risk products, although there is a lack of evidence and there is no FDA authorization for those products as a modified risk tobacco product [35]. These findings broadly support the discovery of quitting-related posts of oral nicotine pouches in our study, as well as the unanticipated conclusion that the majority of these posts have positive sentiments.

Given the relatively short amount of time that oral nicotine pouch products have been on the market, the short- and long-term health effects of oral nicotine pouches use are still unknown and need further investigation. Our investigation of Reddit posts related to oral nicotine pouches identified some health symptoms mentioned by the users such as gum problems, which provided indications of what users may be experiencing and of potential health harms to be investigated. The effects of oral nicotine products on periodontal health and the upper airway of the lung are not known. There is a possibly that these products will cause oral submucous fibrosis [36]. Recent studies indicate that oral nicotine pouches are likely to expose users to lower levels of toxic components compared with snus and other tobacco products [10], though they are still addictive and could be damaging to human health [6]. Although health effects, especially long-term effects, of oral nicotine pouches remain to be determined, we showed in this study that there was an increasing trend of mentioning oral nicotine pouches, and that over half of related Reddit posts had positive attitudes toward oral nicotine pouches. Therefore, potential regulatory policy might be considered to prevent possible health concerns related to oral nicotine pouches.

Limitations

The generalizability of these results is subject to several limitations. First, the paucity of noncommercial posts limited the scope of this study. Due to the fact that oral nicotine pouches are emerging new tobacco products, only 653 posts were included in our analysis during our study period from February 15, 2019, to February 12, 2021.

Thus far, there have been limited analyses of social media data on oral nicotine pouches. Our research provides a first and up-to-date insight into the current public perception and popular topics regarding oral nicotine pouches discussed on Reddit. Although this is a start, further research is needed due to the substantial growth of the oral nicotine pouch market. Our study only discusses the broad health symptoms mentioned in the Reddit posts. Whether those health symptoms are related to use of oral nicotine pouches and the health effects of oral nicotine pouches need to be investigated in future human studies.

Conclusions

Our research provides a first insight into the current public perceptions on Reddit about oral nicotine pouches, which provide valuable information for regulators and policy makers. The trending topics summarized in our study were reflections of what the public was interested in regarding oral nicotine pouches. Our results only provided very preliminary information about the potential health effects related to oral nicotine pouches. Future studies are warranted to evaluate the health effects of oral nicotine pouches for regulation purposes.

Acknowledgments

This study was supported by the WNY Center for Research on Flavored Tobacco Products (CRoFT) under cooperative agreement U54CA228110 from the National Cancer Institute of the National Institutes of Health and the US Food and Drug Administration.

Data Availability Statement

Data are available upon reasonable request.

Authors' Contributions

ZX, RGM, and DL conceived and designed the study. YS, JZ, ZX, and DL analyzed the data. YS, JZ, ZX, RGM, and DL wrote the manuscript. DJO, IR, and SM edited the manuscript. All authors edited and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Flowchart of data collection and preprocessing.
[PDF File (Adobe PDF File), 47 KB - jmir_v24i7e37071_app1.pdf]

Multimedia Appendix 2
Example Reddit posts in each topic and subtopic category with different sentiments.
[DOCX File , 26 KB - jmir_v24i7e37071_app2.docx]

Multimedia Appendix 3
Online collection of oral nicotine pouch brands, products, and flavors.
[DOCX File , 48 KB - jmir_v24i7e37071_app3.docx]

References

1. Andersson M, Lundqvist K. Predicting Tobacco-Free Nicotine Pouch Intention Among Swedish Young Adults: Gender and the Proximal Predictors of the Theory of Triadic Influence. LUP Student Papers. URL: http://lup.lub.lu.se/student-papers/record/9021102 [accessed 2021-07-06]
2. Robichaud MO, Seidenberg AB, Byron MJ. Tobacco companies introduce 'tobacco-free' nicotine pouches. Tob Control 2020 Dec 21;29(e1):e145-e146. [doi: 10.1136/tobaccocontrol-2019-055321] [Medline: 31753961]
3. Plurphanswat N, Hughes JR, Fagerström K, Rodu B. Initial Information on a Novel Nicotine Product. Am J Addict 2020 Jul 16;29(4):279-286. [doi: 10.1111/ajad.13020] [Medline: 32176374]
4. Global Nicotine Pouches Market Research Report 2021. QYResearch Group. URL: https://www.marketresearch.com/QYResearch-Group-v3531/Global-Nicotine-Pouches-Research-14261122/ [accessed 2021-07-06]
5. Smokefree products in the US. SwedishMatch. URL: https://www.swedishmatch.com/Our-business/smokefree/Market-development/smokefree-products-in-the-us/ [accessed 2021-07-06]
6. Stanfill S, Tran H, Tyx R, Fernandez C, Zhu W, Marynak K, et al. Characterization of total and unprotonated (free) nicotine content of nicotine pouch products. Nicotine Tob Res 2021 Aug 18;23(9):1590-1596. [doi: 10.1093/ntr/ntab030] [Medline: 34233354]
7. Harrell M, Weaver S, Loukas A, Creamer M, Marti C, Jackson C, et al. Flavored e-cigarette use: Characterizing youth, young adult, and adult users. Prev Med Rep 2017 Mar;5:33-40 [FREE Full text] [doi: 10.1016/j.pmedr.2016.11.001] [Medline: 27896041]
8. Morean ME, Butler ER, Bold KW, Kong G, Camenga DR, Cavallo DA, et al. Preferring more e-cigarette flavors is associated with e-cigarette use frequency among adolescents but not adults. PLoS One 2018;13(1):e0189015 [FREE Full text] [doi: 10.1371/journal.pone.0189015] [Medline: 29300749]
9. Gravely S, Cummings K, Hammond D, Lindblom E, Smith DM, Martin N, et al. The association of e-cigarette flavors with satisfaction, enjoyment, and trying to quit or stay abstinent from smoking among regular adult vapers from Canada and the United States: Findings from the 2018 ITC four country smoking and vaping survey. Nicotine Tob Res 2020 Oct 08;22(10):1831-1841 [FREE Full text] [doi: 10.1093/ntr/ntaa095] [Medline: 32449933]
10. Azzopardi D, Liu C, Murphy J. Chemical characterization of tobacco-free "modern" oral nicotine pouches and their position on the toxicant and risk continuums. Drug Chem Toxicol 2021 May 25:1-9. [doi: 10.1080/01480545.2021.1925691] [Medline: 34034614]
11. Lawler TS, Stanfill SB, Zhang L, Ashley DL, Watson CH. Chemical characterization of domestic oral tobacco products: total nicotine, pH, unprotonated nicotine and tobacco-specific N-nitrosamines. Food Chem Toxicol 2013 Jul;57:380-386. [doi: 10.1016/j.fct.2013.03.011] [Medline: 23517910]
12. Patel S. Reddit claims 52 million daily users, revealing a key figure for social-media platforms. The Wall Street Journal. 2020. URL: https://www.wsj.com/articles/reddit-claims-52-million-daily-users-revealing-a-key-figure-for-social-media-platforms-11606822200 [accessed 2022-05-19]
13. Lu X, Chen L, Yuan J, Luo J, Luo J, Xie Z, et al. User perceptions of different electronic cigarette flavors on social media: Observational study. J Med Internet Res 2020 Jun 24;22(6):e17280 [FREE Full text] [doi: 10.2196/17280] [Medline: 32579123]
14. 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]
15. Directory contents. Pushshift. URL: http://files.pushshift.io/reddit/submissions [accessed 2022-06-23]
16. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x]
17. Graneheim UH, Lindgren B, Lundman B. Methodological challenges in qualitative content analysis: A discussion paper. Nurse Educ Today 2017 Sep;56:29-34. [doi: 10.1016/j.nedt.2017.06.002] [Medline: 28651100]
18. Zou C, Wang X, Xie Z, Li D. Perceptions of the IQOS heated tobacco product on Twitter in the United States. Front Commun 2021 Aug 20;6:1. [doi: 10.3389/fcomm.2021.728604]

19. Gao Y, Xie Z, Li D. Electronic cigarette users' perspective on the COVID-19 pandemic: observational study using Twitter data. JMIR Public Health Surveill 2021 Jan 05;7(1):e24859 [FREE Full text] [doi: 10.2196/24859] [Medline: 33347422]
20. Tao C, Diaz D, Xie Z, Chen L, Li D, O'Connor R. Potential impact of a paper about COVID-19 and smoking on Twitter users' attitudes toward smoking: observational study. JMIR Form Res 2021 Jun 15;5(6):e25010 [FREE Full text] [doi: 10.2196/25010] [Medline: 33939624]
21. Lu X, Sun L, Xie Z, Li D. Perception of the Food and Drug Administration electronic cigarette flavor enforcement policy on Twitter: observational study. JMIR Public Health Surveill 2022 Mar 29;8(3):e25697 [FREE Full text] [doi: 10.2196/25697] [Medline: 35348461]
22. Sun L, Lu X, Xie Z, Li D. Public reactions to the New York State policy on flavored electronic cigarettes on Twitter: observational study. JMIR Public Health Surveill 2022 Feb 03;8(2):e25216 [FREE Full text] [doi: 10.2196/25216] [Medline: 35113035]
23. Prokop H. Nicotine pouches: tobacco's fast-growing subcategory. CSP Daily News. URL: https://www.cspdailynews.com/tobacco/nicotine-pouches-tobaccos-fast-growing-subcategory [accessed 2022-06-16]
24. Tobacco products global market report 2021: COVID-19 impact and recovery to 2030. The Business Research Company. URL: https://www.reportlinker.com/p06018952/Tobacco-Products-Global-Market-Report-COVID-19-Impact-and-Recovery-to.html?utm_source=GNW [accessed 2022-06-16]
25. What are communities or 'subreddits'? Reddit. URL: https://reddit.zendesk.com/hc/en-us/articles/204533569-What-are-communities-or-subreddits- [accessed 2021-07-06]
26. O'Connor RJ, Norton KJ, Bansal-Travers M, Mahoney MC, Cummings KM, Borland R. US smokers' reactions to a brief trial of oral nicotine products. Harm Reduct J 2011 Jan 10;8(1):1 [FREE Full text] [doi: 10.1186/1477-7517-8-1] [Medline: 21219609]
27. Struik L, Yang Y. e-Cigarette cessation: content analysis of a quit vaping community on Reddit. J Med Internet Res 2021 Oct 25;23(10):e28303 [FREE Full text] [doi: 10.2196/28303] [Medline: 34694229]
28. Craver R. Reynolds quietly removes nicotine-replacement therapy product from U.S. retail. Winston-Salem Journal. 2019. URL: https://tinyurl.com/rt7pjfm [accessed 2022-05-19]
29. BAT Completes Acquisition of Reynolds. British American Tobacco. 2017 Jul 25. URL: https://www.bat.com/group/sites/UK__9D9KCY.nsf/vwPagesWebLive/DOAPKCXS [accessed 2022-05-19]
30. British American Tobacco. 2021. URL: https://www.bat-science.com/sciencereport [accessed 2022-06-22]
31. Reduced-risk products. British American Tobacco. 2022. URL: https://www.bat.com/provingreducedrisk [accessed 2022-06-22]
32. BAT completes PMTA submissions for the broadest portfolio in the U.S. British American Tobacco. 2020 Sep 09. URL: https://www.bat.com/group/sites/UK__9D9KCY.nsf/vwPagesWebLive/DOBT5R3K [accessed 2022-05-19]
33. Kostygina G, England L, Ling P. New product marketing blurs the line between nicotine replacement therapy and smokeless tobacco products. Am J Public Health 2016 Jul;106(7):1219-1222. [doi: 10.2105/ajph.2016.303057]
34. Gong M, Dunbar MS, Setodji C, Shadel WG. Zonnic: a new player in an old field. Subst Abuse Treat Prev Policy 2017 Sep 06;12(1):40 [FREE Full text] [doi: 10.1186/s13011-017-0124-3] [Medline: 28877727]
35. Czaplicki L, Patel M, Rahman B, Yoon S, Schillo B, Rose SW. Oral nicotine marketing claims in direct-mail advertising. Tob Control 2021 May 06:tobaccocontrol-2020-056446 [FREE Full text] [doi: 10.1136/tobaccocontrol-2020-056446] [Medline: 33958422]
36. Javed F, Kellesarian S, Sundar I, Romanos G, Rahman I. Recent updates on electronic cigarette aerosol and inhaled nicotine effects on periodontal and pulmonary tissues. Oral Dis 2017 Nov 30;23(8):1052-1057. [doi: 10.1111/odi.12652] [Medline: 28168771]

Abbreviations

FDA: Food and Drug Administration

Edited by A Mavragani; submitted 05.02.22; peer-reviewed by G Kong, L Struik, K DeLaughter; comments to author 25.03.22; revised version received 19.05.22; accepted 31.05.22; published 15.07.22.

Please cite as:
Shao Y, Zou J, Xie Z, Mayne RG, Ossip DJ, Rahman I, McIntosh S, Li D
Perceptions of Oral Nicotine Pouches on Reddit: Observational Study
J Med Internet Res 2022;24(7):e37071
URL: https://www.jmir.org/2022/7/e37071
doi:10.2196/37071
PMID:35838764

Original Paper

Tweets Related to Motivation and Physical Activity for Obesity-Related Behavior Change: Descriptive Analysis

Albert Park[1], PhD

Department of Software and Information Systems, College of Computing and Informatics, University of North Carolina-Charlotte, Charlotte, NC, United States

Corresponding Author:
Albert Park, PhD
Department of Software and Information Systems
College of Computing and Informatics
University of North Carolina-Charlotte
Woodward 310H
9201 University City Blvd
Charlotte, NC, 28223
United States
Phone: 1 704 687 8668
Fax: 1 704 687 1652
Email: al.park@uncc.edu

Abstract

Background: Obesity is one of the greatest modern public health problems, due to the associated health and economic consequences. Decreased physical activity is one of the main societal changes driving the current obesity pandemic.

Objective: Our goals are to fill a gap in the literature and study whether users organically utilize a social media platform, Twitter, for providing motivation. We examine the topics of messages and social network structures on Twitter. We discuss social media's potential for providing peer support and then draw insights to inform the development of interventions for long-term health-related behavior change.

Methods: We examined motivational messages related to physical activity on Twitter. First, we collected tweets related to physical activity. Second, we analyzed them using (1) a lexicon-based approach to extract and characterize motivation-related tweets, (2) a thematic analysis to examine common themes in retweets, and (3) topic models to understand prevalent factors concerning motivation and physical activity on Twitter. Third, we created 2 social networks to investigate organically arising peer-support network structures for sustaining physical activity and to form a deeper understanding of the feasibility of these networks in a real-world context.

Results: We collected over 1.5 million physical activity–related tweets posted from August 30 to November 6, 2018. A relatively small percentage of the tweets mentioned the term *motivation*; many of these were made on Mondays or during morning or late morning hours. The analysis of retweets showed that the following three themes were commonly conveyed on the platform: (1) using a number of different types of motivation (self, process, consolation, mental, or quotes), (2) promoting individuals or groups, and (3) sharing or requesting information. Topic models revealed that many of these users were weightlifters or people trying to lose weight. Twitter users also naturally forged relations, even though 98.12% (2824/2878) of these users were in different physical locations.

Conclusions: This study fills a knowledge gap on how individuals organically use social media to encourage and sustain physical activity. Elements related to peer support are found in the organic use of social media. Our findings suggest that geographical location is less important for providing peer support as long as the support provides motivation, despite users having few factors in common (eg, the weather) affecting their physical activity. This presents a unique opportunity to identify successful motivation-providing peer support groups in a large user base. However, further research on the effects in a real-world context, as well as additional design and usability features for improving user engagement, are warranted to develop a successful intervention counteracting the current obesity pandemic. This is especially important for young adults, the main user group for social media, as they develop lasting health-related behaviors.

(J Med Internet Res 2022;24(7):e15055) doi:10.2196/15055

KEYWORDS

obesity; motivation; exercise; peer support; social network analysis; social computing; consumer health information; informatics; information science; social support; communications media

Introduction

Obesity is continually increasing around the world [1]. The annual economic cost of obesity is a substantial burden [2], because it is associated with type 2 diabetes, heart disease, stroke, arthritis, high blood pressure, and various cancers [1,3]. For example, total obesity-related medical expenditures are estimated at $116 billion in the United States [4]. Despite obesity's serious health and economic consequences, two-thirds of US adults and nearly a third of children aged 6 to 19 years are either at risk for obesity or are obese [3,5]. Society-level behavior changes, including reduced physical activity, have been determined to be among the main contributors driving the current obesity pandemic [6]. To counteract this modern public health problem, we not only need to initiate society-level behavior changes (eg, increase physical activity), but also need to ensure adherence to the changed behaviors [7].

Studies have demonstrated long-term (ie, more than 18 months) exercise adherence challenges for obese individuals [8], even for individuals who had near-death experiences [9,10], due to issues such as dropout [11]. Peer support has been successful in sustaining changed behaviors, especially in the context of the Alcoholics Anonymous group [12,13]. Peer support has also been suggested to be more effective than financial incentives [14] or nurse care management [15] for people with diabetes. The concept has been successfully applied in mental health [16], acute substance abuse [17], diabetes control [14,15], and efforts to maintain physical activity [18]. However, not all peer support systems are successful [8]. One major contributing factor for developing an effective peer support system is to identify proper peer support pairs [19] or workers [16].

Social media has become an integral part of daily life in the United States [20]. Individuals use social media to freely express their thoughts and to interact with geographically dispersed like-minded individuals. Given the increasing use of social media, health-related research utilizing these platforms is also increasing. For instance, social media data have been applied in improving public health outcomes [21], conducting disease surveillance [22-24], enhancing personal health care experiences [19,25,26], examining the daily struggles of living with mental health conditions [27], predicting adverse drug reactions [28], and understanding user responses during a global pandemic [29]. Despite evidence suggesting that social media has the potential to influence individuals [30-37] and to identify exercise partners [38], we lack knowledge on how users naturally utilize social media for peer support, especially how social media is used to provide motivation for exercise adherence in the context of obesity and weight loss.

Many studies have examined social media or incorporated it into their design to investigate attempts to change health-related behaviors [30], such as by increasing physical activity [32-34], losing weight [35], and adhering to exercise programs [36,37]. However, substantial differences have been demonstrated in the use of social media, and these studies have not delivered long-term impacts in reducing the prevalence of obesity. Moreover, effective design elements are unknown; thus, many studies primarily focus on designing weight loss or weight management interventions that are delivered via social media [39-47], or they investigate relevant elements for higher user engagement [48]. To develop an effective intervention, it is imperative to first understand how users are organically utilizing the platform to provide or gain motivation and peer support, defined here as the exchange of motivational tweets for exercise in general, and then draw insights to inform the development of an intervention that provides peer support for long-term exercise adherence. In this study, we characterize the naturally occurring social network structures of peer support pairs or groups for motivating physical activity via a mixed methods approach. More specifically, we examine messages related to physical activity and motivation on a social media platform (Twitter). This study seeks to answer 2 research questions (RQs) to fill the gap in the literature. RQ1: Do Twitter users utilize the platform to provide or gain motivation for physical activity? RQ2: Do individuals use Twitter to form peer support pairs or groups for motivating physical activity? We restricted our analysis to publicly available discussion content.

Methods

Data Collection and Social Media Site

Twitter is a popular, free-to-use microblogging social media platform on which users can instantly broadcast short messages. These short messages are called tweets. Tweets are limited to 280 characters (this changed from 140 characters in November 2017) and are meant to be broadcasted to the world. Twitter users, however, can also send out directed, conversational messages by using the @ symbol followed by a Twitter user's ID. Twitter does not have topically focused groups, but Twitter users can use the # symbol to label and categorize their tweets with searchable keywords. We employed the Python library Tweepy [49] and used hashtags from a previous study [50], including #weightloss, #diet, #fitness, and #health, to collect tweets related to physical activity and weight loss. We then extracted motivation-related tweets to answer RQ1 and reciprocal conversational interactions to answer RQ2. In this study, we only examined tweets in English.

RQ1: Do Twitter Users Utilize the Platform to Provide or Gain Motivation for Physical Activity?

We first used a lexicon-based approach, which is a method to identify related text passages based on a set of words or phrases, to extract and examine motivation-related tweets from the overall physical activity tweets. A number of different motivation-related terms were considered, but for this study, we only used a single term, *motivation*, for the following three reasons. First, Twitter users frequently use acronyms and abbreviations due to the 280-character length limitation. However, in an online environment, these terms have been

shown to change over time [51]. The same acronyms and abbreviations can also be used differently in different communities, which have specific cultures [52]. Second, we placed a higher priority on precision than recall in order to accurately observe user behavior and reduce noise. Third, many popular motivation-related hashtags contain the term *motivation* (eg, #fitnessmotivation and #mondaymotivation). Thus, we filtered for tweets that contained the keyword *motivation*. Specifically, we first preprocessed the entire data set to convert the text to lowercase. Then, to extract tweets containing our key terms from the entire data set, we employed a lexicon-based approach and extracted tweets and data from their "user" and "entities" fields, including user names, screen names, locations, timestamps, and user mentions. We extracted and included any partial matches in this process to cover a wide variation in terms, such as #mondaymotivation.

To understand temporal trends in motivation-related messages, we first analyzed the data by week and by time of day. In this analysis, we included data from between September 1, 2018, and November 2, 2018, because the tweets in our original data had an uneven number of days of the week and were collected at different starting and ending time points. The hours of the day were grouped into eight periods: (1) early morning (4:00 AM to 6:59 AM), (2) morning (7:00 AM to 9:59 AM), (3) late morning (10:00 AM to 12:59 PM), (4) afternoon (1:00 PM to 3:59 PM), (5) evening (4:00 PM to 6:59 PM), (6) late evening (7:00 PM to 9:59 PM), (7) night (10:00 PM to 12:59 AM), and (8) late night (1:00 AM to 3:59 AM).

To characterize and understand these messages, we first qualitatively examined retweets—reposted or forwarded tweets—related to motivation. Retweets are important in Twitter, because they have a higher chance of reaching a large number of users and can represent messages that embody the collective ideation of the platform [53]. Thus, we qualitatively analyzed the general themes of retweeted tweets following the thematic analysis process [54,55]. A prespecified threshold of at least 50 retweets was applied for a tweet to be analyzed, due to the size of the data set. All of the quotes in this paper have been deidentified and slightly modified to protect the privacy of users following the guidelines suggested by a previous paper [56].

To quantitatively expand the retweet analysis, we created a topic model using latent Dirichlet allocation (LDA) [57] to identify main topics from the larger data set. Topic modeling is an effective method to discover topics in a large data set and can identify a set of topics from a given set of documents based on document-level word co-occurrences. In our study, each tweet was considered a single document. First, we preprocessed the data set by removing URLs and tweets with fewer than 5 words. Similar exclusion criteria were applied to reduce noise in a previous study that used social media data [58]. Then, nouns were extracted using the Python natural language toolkit library [59] and used in the topic modeling process. LDA requires a predetermined number of topics, which can be a limitation. After experimenting with varying numbers of topics, we chose 10 topics related to motivation and physical activity on Twitter. An LDA topic model was generated by the Python library genism [60], followed by manual examination and labeling of the identified topics. To visually review these topics and occurrences of associated terms, we then visualized the topic model (ie, the main topics and their top 50 associated words) as a word cloud using the Python library wordcloud (Multimedia Appendix 1) [61]. Word cloud visualization is simple and user-friendly, yet scalable and preferred by users [62,63].

RQ2: What Types of Reciprocal Conversational Social Network Structures Are Formed on Twitter for Motivating Physical Activity?

We generated 2 social network visualizations to verify whether social media enables organically formed peer support pairs or groups for motivating physical activity and to examine their social network structure. The first Twitter social network visualization represents a social structure of reciprocal, conversational interactions that are captured when users direct their tweets using the @ symbol to other Twitter IDs. These interactions are dyadic ties among Twitter users (determined by Twitter ID), who are the social actors. We first extracted all tweets directed toward other users by checking if the @ symbol was placed in conjunction with a Twitter ID in a tweet. We then use Gephi [64], a popular network visualization tool, to visualize the social network and apply the OpenOrd [65] layout to gain an overview of the network structure. The edge weight is determined by the number of interactions among Twitter users.

To examine Twitter's real-world potential for connecting physical activity partners [38], we generated a second social network visualization of users who disclosed their location information. The purpose of this social network visualization was to visually examine the prevalence of social interactions that are generated from the same physical location and to assess the relationship between users' current practices and the platform's potential utility for connecting physical activity partners. Twitter users can opt to share their location information in their user profile. The Twitter API also provides a mechanism to track the geocode of a tweet; however, a previous study suggests that only 1.5% of tweets are geotagged [66]. Thus, we extracted users' location information from their profile for the second social network visualization and also recorded the number of geotagged tweets and users who opted to disclose this information. We removed social actors without location information, and then estimated peer-supporting physical activity partners (ie, socially and physically connected users). We denoted this information using a coloring schema in the second social network visualization. We used the Fruchterman Reingold graph layout algorithm [67] for the second visualization, which had a network structure with considerably fewer social actors and dyadic ties.

Ethics Statement

The study was determined to not have human subjects by the University of North Carolina-Charlotte's Institutional Review Board (Ethics Committee).

Results

Data Set

We collected over 1.5 million physical activity–related tweets made between August 30, 2018, and November 6, 2018. Table

1 summarizes the overall data set and the data sets for each research question.

Table 1. Summary of data set.

Tweet type	Physical activity tweets (N=1,528,439), n (%)	Unique Twitter IDs (N=474,402), n (%)
Motivation (research question 1)	39,703 (2.6)	13,561 (2.86)
Directed motivation (research question 2)	15,720 (1.03)	6558 (1.38)

RQ1: Do Twitter Users Utilize the Platform to Provide or Gain Motivation for Physical Activity?

Overview

Of the 1.5 million physical activity–related tweets, we found that 2.6% (39,703/1,528,439) contained the term *motivation*. This is a relatively small proportion of the overall physical activity–related tweets, but it still amounted to 39,703 tweets made by 13,561 unique Twitter IDs. Of these 39,703 motivational messages, the most were sent out on Mondays (7070 tweets) and during morning (5668 tweets) or late morning (5748 tweets) hours (7:00 AM to 9:59 AM and 10:00 AM to 12:59 PM, respectively). Other days had substantially fewer tweets (from Tuesday to Sunday, in order: 5278, 5487, 5374, 5492, 5819, and 5220 tweets). Similarly, after late morning, motivational messages gradually decreased until the morning of the next day (from afternoon to early morning of the next day, in chronological order: 5247, 5284, 4955, 4930, 3866, and 4042 tweets). Within the morning and late morning hours, the most tweets were sent around lunchtime, from 11:00 AM to 11:59 AM (1987 tweets), followed by the 7:00 AM to 7:59 AM period (1949 tweets). Specific topics related to these tweets are further examined and described via qualitative analysis of retweets and topic modeling of the entire motivation data set in the following section.

In our data set, we found that 13,268 of the 39,703 (33.42%) overall motivation tweets were retweets. These 13,268 retweets retweeted 5915 (44.58%) original tweets. Of these 5915 original tweets, 18 were retweeted more than a prespecified threshold of 50 times. These 18 tweets were retweeted 1562 times, comprising 11.77% of the total number of retweets. Qualitative analysis of these 18 tweets identified the following 3 themes: *motivation*, *promoting*, and *information*.

Providing Motivation (Self, Process, Consolation, Mental, Quotes)

Providing motivation was the most common theme. However, a number of different subthemes were identified within this theme, including *self-motivation*, *motivation for continuing the physical activity journey/process*, *self-consolation*, *mental aspects of motivation*, and *use of quotes*. We categorized 12 of the 18 retweets as motivation. The following quote is a slightly modified example of a motivation retweet: “Finding motivation to workout doesn’t need to be a difficult process....”

Promoting (Individuals or Groups)

We categorized 4 of the 18 retweets into the *promoting* theme. Two types of promoting were identified, *promoting individuals* and *promoting groups*, for physical activity–related matters. One promoting method was to tweet out the names of individuals; thus, we do not show those tweets. Another promoting method was done using the following format in our data set: “Follow us for...”

Information (Sharing or Requesting Information)

We categorized 2 of the 18 retweets into the *information* theme. In these retweets, Twitter users voluntarily shared health tips or health-related information. It is also important to note that some retweets could have been made to promote products or brands, a practice known as “astroturfing.” Astroturfing is when sponsors or organizations mask themselves as Twitter users to promote their products [68,69]. We categorized these tweets as *information* because of the subtle nature of astroturfing on Twitter and a lack of knowledge regarding the motives for sharing such information. The following quote is a modified example of information sharing: “The Truth About The...”

Our topic modeling analysis expanded the findings of the qualitative analysis of retweets. Topic modeling revealed that the *providing motivation* topic had different types. Users often revealed their personal motivation for physical activities, which included tweets with the topics of *appearance*, *weight loss*, *lifestyle*, *health*, *wealth*, *mental*, and *personal motivation*. We also identified *workout inspiration* and *daily motivation* as subtopics that were part of the *providing motivation* topic.

Topic modeling also identified topics related to sharing or requesting information, specifically information related to the *nutrition-diet* topic. Compared to the retweets in this category, the tweets contained a greater number of products, tips, and information. Additional topics were identified in this process, mainly different types of workouts, including the *muscular training* and *bodybuilding* topics, presumably used by weightlifters. Topic modeling, however, did not identify a topic related to *promoting*. A total of 10 topics were identified via topic modeling; Table 2 shows sample terms from each of the identified topics. Multimedia Appendix 1 summarizes and provides an overview of all the identified topics and their related terms.

Table 2. Ten topics and samples of associated terms found in motivational tweets identified via topic modeling.

Topic label	Selected sample terms
Workout inspiration	motivation, inspiration, fitness, love, goals, fitnessmotivation, workout, quotes, music, mood, determination, fitnessquotes
Appearance	fitbeauty, bikinibody, beautycoolgym, beautycoolmotivation, ass, erossensual, look, amazing, weightloss
Muscular training	focus, workout, exercise, muscle, body, power, coach, hardwork, mindset, grind, strength, nutrition, passion, trainer, strong, bodygoals
Bodybuilding	gym, motivation, bodybuildingmotivation, gymlife, tip, abs, men, bodybuilder, goal, women, protein, squats, physique
Weight loss	weightloss, weightlossmotivation, weightlossjourney, tips, facts, resources, weightlosstransformation, hotbodiedsnaps
Lifestyle	body, health, mind, lifestyle, healthyfood, enjoy, fitnessgirl, gymshark, pain, hope, momlife, bodytransformation, obesity
Nutrition-diet	nutrition, diet, healthy, health, food, fashion, truth, coffee, ketogenic, weightwatchers, travel, science, supplements, foods, deal
Health; wealth; mental	health, fitness, muscle, excuses, gold, pushpullgrind, dreams, goals wealth, mind, healthylife, money, grindout, business
Personal motivation	weight, cardio, bodybuilding, weightloss, fitmotivation, fitnessmodel, fitnessaddict, keto, hormones, strength, girls, beach, girl
Daily motivation	workout, mondaymotivation, fitnessmotivation, day, week, time, darebees, days, Tuesday, work, session, mondaymood, saturdaymotivation

RQ2: What Types of Reciprocal Conversational Social Network Structures Are Formed on Twitter for Motivating Physical Activity?

The overall social network of directed tweets consisted of 14,028 dyadic ties among 10,889 social actors (determined by Twitter ID). It is important to note that some tweets contained multiple indications of where they directed their tweets, included self-mentioning, or were directed to Twitter IDs that were not found in our data set. This resulted in a greater number of social actors than the total number of Twitter IDs but a lesser number of dyadic ties than the total number of tweets shown in Table 1 and Table 3. Table 3 summarizes the directed tweets and the creators of those tweets (determined by Twitter ID) in our data set. About 68% (4459/6558) of the Twitter users disclosed their location in their profile; however, users rarely disclosed their exact location by providing geocode information. Similar to a previous study [66], our data set consisted of about 1% (94/6558) geotagged tweets and Twitter users, resulting in less than 100 users who opted to disclose their exact location.

Figure 1 shows an overview of the Twitter social network in our data set. Twitter users formed hundreds of small groups that sent motivation-related tweets to each other, while a small number of larger groups with a higher number of dyadic ties and social actors were formed. Figure 2 shows a social network of location-based Twitter user interactions. After removing social actors without location information, self-mentioning tweets, and interactions with only one-sided location information, the social network structure consisted of 2878 dyadic ties among 1112 social actors, a smaller number of social actors and dyadic ties than shown in Table 3. Of these 2878 social interactions, 1.88% (54/2878) were made from the same physical location.

Table 3. Summary of directed motivation tweet data set.

Tweet type	Total tweets (N=15,720), n (%)	Unique Twitter IDs (N=6558), n (%)
Directed at Twitter ID with location in the profile ID	10,533 (67)	4459 (67.99)
Directed tweet with geocode	159 (1.01)	94 (1.43)

Figure 1. A social network of Twitter users sending directed tweets.

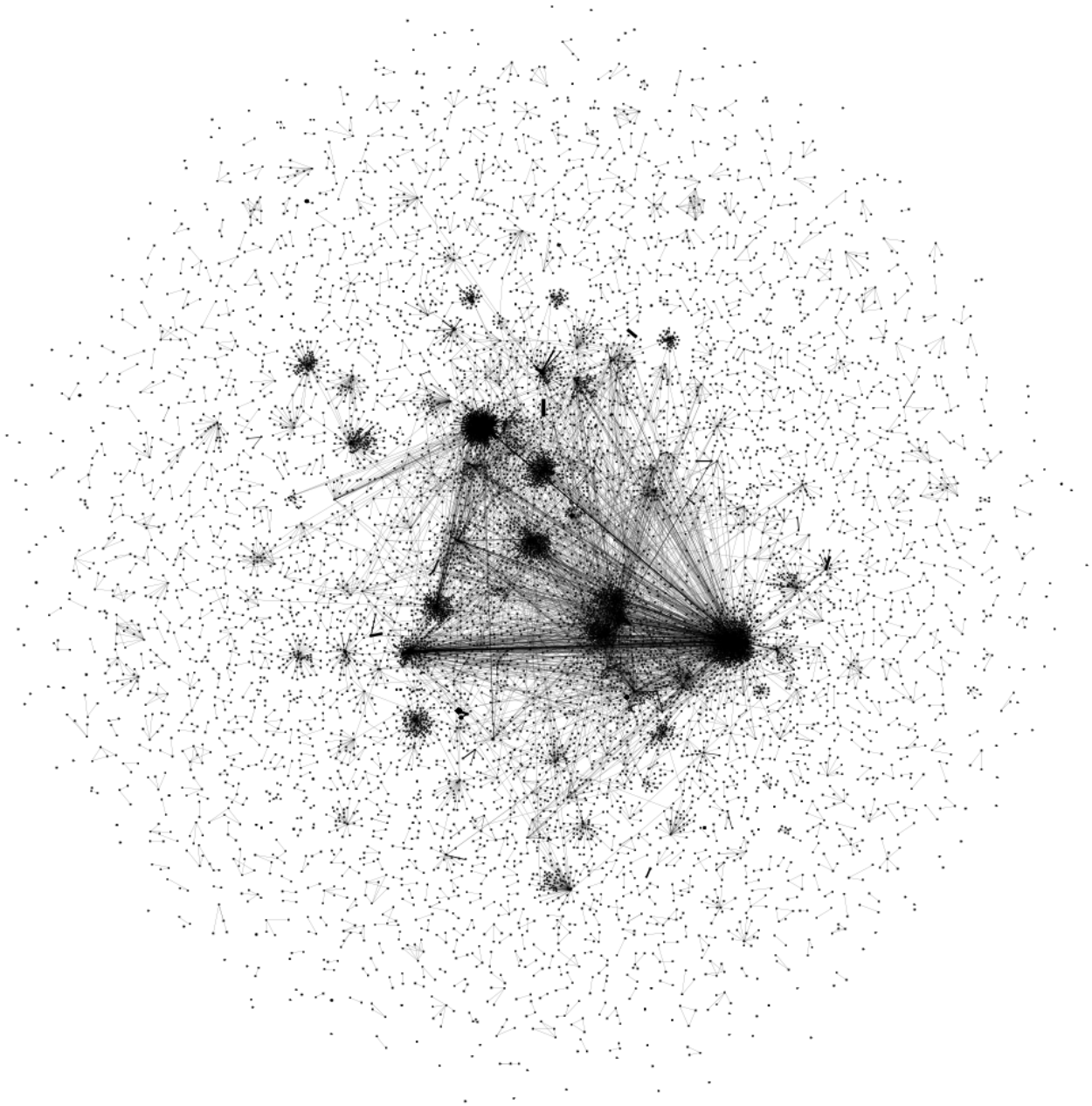
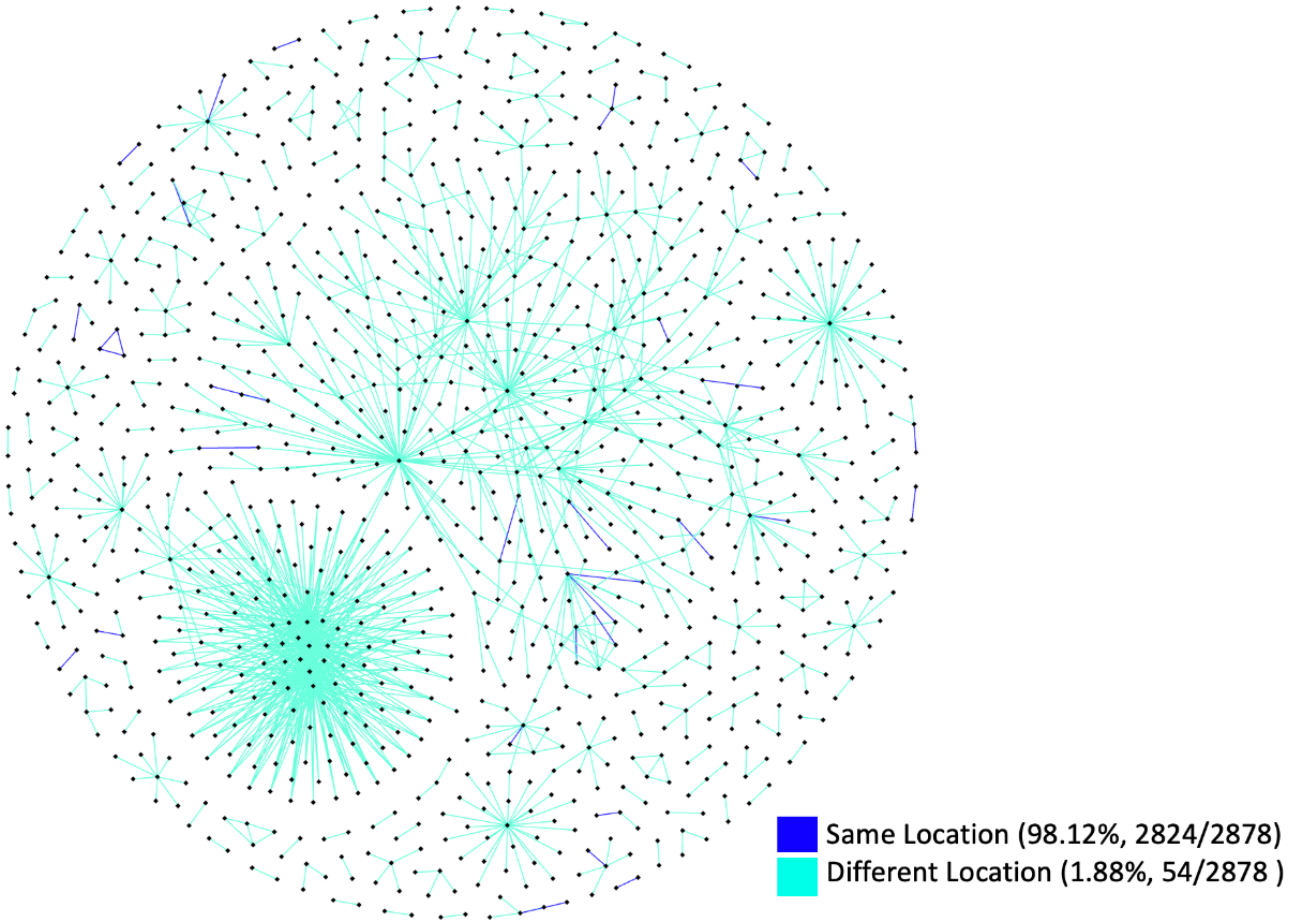


Figure 2. A social network of location-based Twitter user interactions. Twitter users from the same location are denoted in blue, whereas Twitter users from different locations are denoted in green.



Discussion

This paper characterizes how Twitter users organically use a social media platform for motivating physical activity. To better understand Twitter's potential as a long-term intervention platform, we specifically investigated and report elements that were related to peer support and motivation for physical activity. We also discuss the practical implications of our work and future directions.

Practical Implications and Future Directions

Tweeting for Motivation

Twitter users use the platform to discuss and broadcast messages about motivation and physical activity. We extracted and analyzed a large number of motivation-related tweets (39,703 tweets from 13,561 unique Twitter IDs); however, these tweets were made by a small percentage of individuals (13,561/474,402, 2.86%) who tweeted about physical activity. Similarly, motivation-related tweets made up a small percentage (39,703/1,528,439, 2.6%) of overall physical activity–related tweets. In our extraction process, we purposely chose precision over recall, thus potentially reducing the overall number of motivation-related tweets. However, it is apparent that motivation was not the main focus of a large portion of the physical-activity tweets. A future study should identify other topics related to physical activity and examine how to further encourage and increase peer support for positive behavior change in social media. Despite this, messages by users to motivate themselves, as well as others, were typically sent on Mondays and in the morning and late afternoon hours. This could indicate that Twitter users need the most motivational support and exercise during these time periods. A hybrid peer-support intervention augmented with an automatic motivational feature may be able to automatically provide encouraging messages when motivation is most needed (eg, Mondays and during morning and late afternoon hours), if peer supporters cannot. A future study should investigate the effectiveness of a similar hybrid intervention to verify whether automatically generated messages can help sustain and encourage physical activity among users.

Usage

Among retweets, we identified the following 3 themes: *motivation*, *promoting*, and *information*. As individuals promote themselves on social media, they gain more influence over other users. The characteristics of health information provided by credentialed experts and individuals have been suggested to be different [70,71]. Capitalizing on the influence of these individuals could provide another avenue for changing health-related behavior. However, a better understanding of the role of these individuals and their messages is needed. For example, although individuals' experience-based information provides many benefits, it could also be misleading without proper contextual information. Twitter, with its 280-character limit, could mislead many individuals, and intermediary

monitoring of the quality of health information may thus be required in health-behavior interventions that use the Twitter platform. Another interesting finding was that Twitter users use the platform for physical activity self-motivation. The rationale behind this phenomenon is largely unknown, and this could be another mechanism for sustaining physical activity.

Main User Demographics

The topic model revealed new information about the users. For example, our model suggests that a large portion of these individuals were weightlifters (suggested by the *muscular training* and *bodybuilding* topics) or trying to lose weight (suggested by the *weight loss* and *nutrition-diet* topics). Our analysis also suggests that individuals need reminders and inspiration to stay physically active (suggested by the *workout inspiration*, *personalmotivation*, and *daily motivation* topics). Other motivation-related factors for physical activity included lifestyle choices, appearance, health, and wealth.

Other Findings

Previous literature suggests that individuals do not want to share their problems using individually identifying social media accounts, such as on Facebook [72]. Our analysis revealed a few terms that described users' physical ability with negative terminologies (eg, *grind* and *pain*), although they were not prevalent. Future work should further identify struggles and concerns that may be harmful for individuals or that can be improved through better design.

Another theme was sharing information voluntarily. Qualitative analysis of retweets led us to suspect that this theme may have arisen from astroturfing, but further analysis is warranted to confirm this observation. Despite the potential misuse of social media, we also found that Twitter users positively motivate themselves and other users by sharing quotes, comforting themselves and others, and discussing the process of their journey and mental aspects of staying physically active. These findings are design elements that need to be considered for developing a successful intervention using social media to increase physical activity.

Peer Support

Online peer support systems, in which peer-to-peer interactions can scale up to arbitrary values of n to n, hold promise for identifying successful peer support groups. Features for identifying physical activity partners exist on a number of online social health activity networks [38], which suggests the need and preference for peer support concerning physical activity. Forging long-lasting connections among a large user base remains a research challenge, although past work has suggested that written communication style and personality matching are important in the success of online peer support systems [19,73]. Despite these challenges, we found empirical evidence that Twitter users use the platform to motivate individuals from different physical locations (2824/2878, 98.12%) to increase their physical activity. However, we do not know how Twitter users organically created these peer-to-peer connections or whether these individuals already knew each other prior to Twitter communication.

Twitter users forge connections using the @ symbol and directly provide or gain motivation for achieving their physical activity goals. In our social network analysis, we verified the existence of a few large, highly interconnected groups as well as hundreds of small groups with fewer interactions. Investigating linguistic differences in topics and communication style and differences in personality matching between the large and small groups could show individuals' preferences and needs for varying sizes of motivation groups. Because our study provides a snapshot of existing peer-to-peer interactions, we are uncertain how Twitter users form these interactions, find peer support partners, and join groups, and we do not know the long-term effects. Future work could ask Twitter users about their experiences related to forming and continuing peer support systems on Twitter by using surveys and interviews to gain a deeper understanding of long-term behavior changes via social media. Similarly, understanding the effects of strong and weak ties in large groups is important in utilizing social factors for this purpose.

Our findings suggest that Twitter has not been connecting local users and that Twitter users are more engaged with geographically dispersed individuals. Less than 2% (54/2878) of directed tweets were made from the same physical location. Our findings indicate that geographic location is of low importance in providing peer support; as long as the support provides motivation, it can succeed despite the peers having few common factors (eg, the weather) affecting their physical activity. This presents a unique opportunity to identify successful peer support groups that provide motivation in large user bases.

Online peer support systems were observed in our study, but we do not know their effects in a long-term, real-world context, which should be considered in related future work. For example, the effects of online peer support related to (1) fostering positive behavior change and (2) sustaining long-term engagement need to be further examined. A previous study suggests that higher engagement with a group leads to more weight loss [74], but sustaining continuous interaction is a prominent challenge for many online social groups [75-80]. Our analysis of retweets also highlights potential commercial influence, or astroturfing, in Twitter networks. The extent of astroturfing and its influence over social structures and user behavior are unconfirmed and warrant further investigation to control commercial and astroturfing effects.

User Privacy

Although Twitter data are public, and tweets are meant to be broadcasted to the world, we discovered consistent user behaviors aimed at protecting user privacy. For example, we found that almost all users opted out of disclosing geocode information (ie, their exact location), as shown in Table 3. Also, even though the majority of users opted to disclose their location in their profile, many did not provide granular information (the limitation section has further details). Research that uses public social media data, including this study, is typically granted exemption from review by Institutional Review Boards in the United States. User behavior in our study indicates that ethical considerations are still needed, especially with regard to privacy [81,82]. In this paper, we did not present user-identifiable

information (eg, Twitter IDs) and modified the example quotations to protect user anonymity.

Limitations

A relatively short timeframe and a lack of seasonal coverage are limitations of our study. Though we did not examine seasonal trends, it is possible that posting frequency could change with the season. Similarly, previous work has shown that social media usage can vary temporally and spatiotemporally [83]. The time period in our study may also have affected trends in posting. Our study time period included a change of seasons, from late summer to fall, which may have been associated with changes in activity and exercise. Similarly, we manually examined nearly 11.77% (1562/13,268) of the total retweets, but the number of retweeted original tweets in our qualitative analysis was relatively small (18/5915). Although we used a prespecified threshold to select original tweets to be qualitatively examined, investigating a larger number of original tweets could have led to more conclusive findings. Moreover, future studies should examine long-term engagement with content among group members to understand the formation, evolution, and collapse of peer networks for the purpose of encouraging physical activities.

Twitter has a diverse user group compared to many other social media platforms, although it is still used more by young adults [84]. This could have created a bias toward younger Twitter users, allowing us to identify particular topics of interest only to this group. We do not know how our findings would apply to other demographics or social media sites. Similarly, our analysis focused on English tweets, and our findings may not be generalizable to non–English speaking countries. Our findings, however, have significant implications for young adults, the main user group of social media, and the health of future society, as they develop lasting health-related behaviors.

Another selection bias was present in the location-based Twitter user interaction analysis (Figure 2). Although the majority of users opted to disclose their location information in their profile (4459/6558, 67.99%), this analysis was conducted with users who self-selected to disclose this information. In the same analysis, an additional limitation was discovered in our manual validation check. We believe that the different location figure (2824/2878, 98.12%) was inflated due to the granularity of information provided by users. For example, in our data set, we found a few users who had identified their location as “Earth” or “America.” Moreover, some users disclosed their information at the state level (eg, North Carolina) and some users disclosed this information at the city level (eg, Charlotte). Although it is possible that they such users were in fact at the same physical location, we were not able to determine this without additional information. In our study, they were thus treated as being in different locations.

In this study, we used hashtags that were identified in a previous study that collected tweets related to physical activity for the purpose of weight loss. However, we did not incorporate postprocessing to include a wider variation in physical activities. The collected tweets were then processed to extract motivation-related tweets for RQ1. Initial key terms are important in collecting a relevant data set, and different sets of terms could have resulted in different findings, especially related to the prevalence of these tweets. Similarly, for RQ1, we used a single keyword, *motivation*, to identify motivation-related tweets. While this approach can identify related tweets with high precision, we should in future extend these study results to retrieve a greater number of relevant messages by developing a classifier. Moreover, major differences, such as communication style, user demographic, and the platform’s intended purpose, exist among social media platforms. To better characterize the landscape of peer support and motivating messages for physical activity in social media, replicating a similar study on a different popular social media platform, such as Facebook or Reddit, is warranted.

Conclusions

Peer support through social media could be a potential means of encouraging positive behavior changes and sustaining that changed behavior. This study examines whether individuals organically use social media to encourage and sustain physical activity. The organic exchange of motivational messages and peer support are found among Twitter users, although this support is more commonly used by individuals from different physical locations. Our findings suggest that geographical location is less important in providing peer support, as long as the support provides motivation, despite the users having few factors in common affecting their physical activity (eg, the weather). Based on existing interaction patterns, we consider that successful motivation-providing peer support groups or interventions could be delivered through social media. However, further research on the effects in a real-world context, as well as additional design and usability features improving user engagement, are warranted to develop a successful intervention to counteract the current obesity pandemic. This is especially important for young adults, the main user group of social media, as they develop lasting health-related behaviors.

Acknowledgments

This work was supported, in part, by funds provided by the University of North Carolina at Charlotte.

Conflicts of Interest

None declared.

Multimedia Appendix 1
A word cloud overview of topic modeling: visual representation of main topics and frequently occurring associated terms.

[PNG File , 268 KB - jmir_v24i7e15055_app1.png]

References

1. Hruby A, Hu FB. The Epidemiology of Obesity: A Big Picture. Pharmacoeconomics 2015 Jul;33(7):673-689 [FREE Full text] [doi: 10.1007/s40273-014-0243-x] [Medline: 25471927]
2. Tremmel M, Gerdtham U, Nilsson PM, Saha S. Economic Burden of Obesity: A Systematic Literature Review. Int J Environ Res Public Health 2017 Apr 19;14(4):435 [FREE Full text] [doi: 10.3390/ijerph14040435] [Medline: 28422077]
3. Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of obesity in the United States, 2009-2010. NCHS Data Brief 2012 Jan(82):1-8 [FREE Full text] [Medline: 22617494]
4. Wang YC, Pamplin J, Long MW, Ward ZJ, Gortmaker SL, Andreyeva T. Severe Obesity In Adults Cost State Medicaid Programs Nearly $8 Billion In 2013. Health Aff (Millwood) 2015 Nov;34(11):1923-1931. [doi: 10.1377/hlthaff.2015.0633] [Medline: 26526251]
5. Hedley AA, Ogden CL, Johnson CL, Carroll MD, Curtin LR, Flegal KM. Prevalence of overweight and obesity among US children, adolescents, and adults, 1999-2002. JAMA 2004 Jun 16;291(23):2847-2850. [doi: 10.1001/jama.291.23.2847] [Medline: 15199035]
6. Hill JO, Peters JC. Environmental contributions to the obesity epidemic. Science 1998 May 29;280(5368):1371-1374. [doi: 10.1126/science.280.5368.1371] [Medline: 9603719]
7. Ory MG, Lee Smith M, Mier N, Wernicke MM. The science of sustaining health behavior change: the health maintenance consortium. Am J Health Behav 2010;34(6):647-659 [FREE Full text] [doi: 10.5993/ajhb.34.6.2] [Medline: 20604691]
8. Perri MG, McAdoo WG, McAllister DA, Lauer JB, Jordan RC, Yancey DZ, et al. Effects of peer support and therapist contact on long-term weight loss. J Consult Clin Psychol 1987 Aug;55(4):615-617. [doi: 10.1037/0022-006X.55.4.615] [Medline: 3624623]
9. van Elderen-van Kemenade T, Maes S, van den Broek Y. Effects of a health education programme with telephone follow-up during cardiac rehabilitation. Br J Clin Psychol 1994 Sep;33(3):367-378. [doi: 10.1111/j.2044-8260.1994.tb01133.x] [Medline: 7994223]
10. Bock BC, Albrecht AE, Traficante RM, Clark MM, Pinto BM, Tilkemeier P, et al. Predictors of exercise adherence following participation in a cardiac rehabilitation program. Int J Behav Med 1997;4(1):60-75. [doi: 10.1207/s15327558ijbm0401_4] [Medline: 16250742]
11. Kelders SM, Van Gemert-Pijnen JEWC, Werkman A, Nijland N, Seydel ER. Effectiveness of a Web-based intervention aimed at healthy dietary and physical activity behavior: a randomized controlled trial about users and usage. J Med Internet Res 2011 Apr 14;13(2):e32 [FREE Full text] [doi: 10.2196/jmir.1624] [Medline: 21493191]
12. Tonigan JS, Toscova R, Miller WR. Meta-analysis of the literature on Alcoholics Anonymous: sample and study characteristics moderate findings. J Stud Alcohol 1996 Jan;57(1):65-72. [doi: 10.15288/jsa.1996.57.65] [Medline: 8747503]
13. Tonigan JS, Miller WR, Connors GJ. Project MATCH Client Impressions About Alcoholics Anonymous. Alcohol Treat Q 2008 Sep 23;18(1):25-41. [doi: 10.1300/J020v18n01_02]
14. Long JA, Jahnle EC, Richardson DM, Loewenstein G, Volpp KG. Peer mentoring and financial incentives to improve glucose control in African American veterans: a randomized trial. Ann Intern Med 2012 Mar 20;156(6):416-424 [FREE Full text] [doi: 10.7326/0003-4819-156-6-201203200-00004] [Medline: 22431674]
15. Heisler M, Vijan S, Makki F, Piette JD. Diabetes control with reciprocal peer support versus nurse care management: a randomized trial. Ann Intern Med 2010 Oct 19;153(8):507-515 [FREE Full text] [doi: 10.7326/0003-4819-153-8-201010190-00007] [Medline: 20956707]
16. Repper J, Carter T. A review of the literature on peer support in mental health services. J Ment Health 2011 Aug;20(4):392-411. [doi: 10.3109/09638237.2011.583947] [Medline: 21770786]
17. Galanter M, Dermatis H, Egelko S, De Leon G. Homelessness and mental illness in a professional- and peer-led cocaine treatment clinic. Psychiatr Serv 1998 Apr;49(4):533-535. [doi: 10.1176/ps.49.4.533] [Medline: 9550249]
18. Davison KK, Jago R. Change in parent and peer support across ages 9 to 15 yr and adolescent girls' physical activity. Med Sci Sports Exerc 2009 Sep;41(9):1816-1825 [FREE Full text] [doi: 10.1249/MSS.0b013e3181a278e2] [Medline: 19657287]
19. Hartzler AL, Taylor MN, Park A, Griffiths T, Backonja U, McDonald DW, et al. Leveraging cues from person-generated health data for peer matching in online communities. J Am Med Inform Assoc 2016 May 05;23(3):496-507 [FREE Full text] [doi: 10.1093/jamia/ocv175] [Medline: 26911825]
20. Perrin A, Atske S. About three-in-ten U.S. adults say they are ‘almost constantly’ online. Pew Research Center. 2021 Mar 26. URL: https://www.pewresearch.org/fact-tank/2021/03/26/about-three-in-ten-u-s-adults-say-they-are-almost-constantly-online; Archived at: https://web.archive.org/web/20220703224341/https://www.pewresearch.org/fact-tank/2021/03/26/about-three-in-ten-u-s-adults-say-they-are-almost-constantly-online [accessed 2022-07-13]
21. Dredze M. How Social Media Will Change Public Health. IEEE Intell Syst 2012 Jul;27(4):81-84 [FREE Full text] [doi: 10.1109/MIS.2012.76]
22. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection--harnessing the Web for public health surveillance. N Engl J Med 2009 May 21;360(21):2153-5, 2157 [FREE Full text] [doi: 10.1056/NEJMp0900702] [Medline: 19423867]

23. Park A, Conway M. Tracking Health Related Discussions on Reddit for Public Health Applications. AMIA Annu Symp Proc 2017;2017:1362-1371 [FREE Full text] [Medline: 29854205]
24. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
25. Friedman C, Elhadad N. Natural Language Processing in Health Care and Biomedicine. In: Shortliffe E, Cimino J, editors. Biomedical Informatics. London, UK: Springer; 2014:255-284.
26. Hartzler AL, McDonald DW, Park A, Huh J, Weaver C, Pratt W. Evaluating health interest profiles extracted from patient-generated data. AMIA Annu Symp Proc 2014;2014:626-635 [FREE Full text] [Medline: 25954368]
27. Park A, Conway M, Chen AT. Examining Thematic Similarity, Difference, and Membership in Three Online Mental Health Communities from Reddit: A Text Mining and Visualization Approach. Comput Human Behav 2018 Jan;78:98-112 [FREE Full text] [doi: 10.1016/j.chb.2017.09.001] [Medline: 29456286]
28. Chee BW, Berlin R, Schatz B. Predicting adverse drug events from personal health messages. AMIA Annu Symp Proc 2011;2011:217-226 [FREE Full text] [Medline: 22195073]
29. Kwon S, Park A. Understanding user responses to the COVID-19 pandemic on Twitter from a terror management theory perspective: Cultural differences among the US, UK and India. Comput Human Behav 2022 Mar;128:107087 [FREE Full text] [doi: 10.1016/j.chb.2021.107087] [Medline: 34744298]
30. Laranjo L, Arguel A, Neves AL, Gallagher AM, Kaplan R, Mortimer N, et al. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. J Am Med Inform Assoc 2015 Jan;22(1):243-256 [FREE Full text] [doi: 10.1136/amiajnl-2014-002841] [Medline: 25005606]
31. Park A, Conway M. Longitudinal Changes in Psychological States in Online Health Community Members: Understanding the Long-Term Effects of Participating in an Online Depression Community. J Med Internet Res 2017 Mar 20;19(3):e71 [FREE Full text] [doi: 10.2196/jmir.6826] [Medline: 28320692]
32. Davies CA, Spence JC, Vandelanotte C, Caperchione CM, Mummery WK. Meta-analysis of internet-delivered interventions to increase physical activity levels. Int J Behav Nutr Phys Act 2012 Apr 30;9:52 [FREE Full text] [doi: 10.1186/1479-5868-9-52] [Medline: 22546283]
33. Napolitano MA, Fotheringham M, Tate D, Sciamanna C, Leslie E, Owen N, et al. Evaluation of an internet-based physical activity intervention: a preliminary investigation. Ann Behav Med 2003;25(2):92-99. [doi: 10.1207/S15324796ABM2502_04] [Medline: 12704010]
34. Vandelanotte C, Spathonis KM, Eakin EG, Owen N. Website-delivered physical activity interventions a review of the literature. Am J Prev Med 2007 Jul;33(1):54-64. [doi: 10.1016/j.amepre.2007.02.041] [Medline: 17572313]
35. Beleigoli AM, Andrade AQ, Cançado AG, Paulo MN, Diniz MDFH, Ribeiro AL. Web-Based Digital Health Interventions for Weight Loss and Lifestyle Habit Changes in Overweight and Obese Adults: Systematic Review and Meta-Analysis. J Med Internet Res 2019 Jan 08;21(1):e298 [FREE Full text] [doi: 10.2196/jmir.9609] [Medline: 30622090]
36. Watson A, Bickmore T, Cange A, Kulshreshtha A, Kvedar J. An internet-based virtual coach to promote physical activity adherence in overweight adults: randomized controlled trial. J Med Internet Res 2012 Jan 26;14(1):e1 [FREE Full text] [doi: 10.2196/jmir.1629] [Medline: 22281837]
37. Richardson CR, Buis LR, Janney AW, Goodrich DE, Sen A, Hess ML, et al. An online community improves adherence in an internet-mediated walking program. Part 1: results of a randomized controlled trial. J Med Internet Res 2010 Dec 17;12(4):e71 [FREE Full text] [doi: 10.2196/jmir.1338] [Medline: 21169160]
38. Nakhasi A, Shen AX, Passarella RJ, Appel LJ, Anderson CA. Online social networks that connect users to physical activity partners: a review and descriptive analysis. J Med Internet Res 2014 Jun 16;16(6):e153 [FREE Full text] [doi: 10.2196/jmir.2674] [Medline: 24936569]
39. Baghaei N, Freyne J, Kimani S, Smith G, Berkovsky S, Bhandari D, et al. SOFA: An Online Social Network for Engaging and Motivating Families to Adopt a Healthy Lifestyle. In: OZCHI '09: Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7. 2009 Presented at: OZCHI '09: Proceedings of the 21st conference of the computer-human interaction special interest group of Australia on Computer-human interaction: design; Nov 23-27, 2009; Melbourne, Australia p. 269-272. [doi: 10.1145/1738826.1738871]
40. Jansen A, Van Mechelen M, Slegers K. Personas and Behavioral Theories: A Case Study Using Self-Determination Theory to Construct Overweight Personas. In: CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI '17: CHI Conference on Human Factors in Computing Systems; May 6-11, 2017; Denver, CO p. 2127-2136. [doi: 10.1145/3025453.3026003]
41. Colineau N, Paris C, Marendy P, Bhandari D, Shu Y. Supporting family engagement in weight management. In: CHI EA '09: CHI '09 Extended Abstracts on Human Factors in Computing Systems. 2009 Presented at: CHI '09: CHI Conference on Human Factors in Computing Systems; Apr 4-9, 2009; Boston, MA p. 3991-3996. [doi: 10.1145/1520340.1520606]
42. Chuah M, Jakes G, Qin Z. WiFiTreasureHunt: a mobile social application for staying active physically. In: UbiComp '12: Proceedings of the 2012 ACM Conference on Ubiquitous Computing. 2012 Presented at: Ubicomp '12: The 2012 ACM Conference on Ubiquitous Computing; Sep 5-8, 2012; Pittsburgh, PA p. 631-632. [doi: 10.1145/2370216.2370339]
43. Kim KK, Logan HC, Young E, Sabee CM. Youth-centered design and usage results of the iN Touch mobile self-management program for overweight/obesity. Pers Ubiquit Comput 2014 Aug 14;19(1):59-68. [doi: 10.1007/s00779-014-0808-x]

44. Patrick K, Marshall SJ, Davila EP, Kolodziejczyk JK, Fowler JH, Calfas KJ, et al. Design and implementation of a randomized controlled social and mobile weight loss trial for young adults (project SMART). Contemp Clin Trials 2014 Jan;37(1):10-18 [FREE Full text] [doi: 10.1016/j.cct.2013.11.001] [Medline: 24215774]
45. Chuah M, Sample S. Fitness Tour: a mobile application for combating obesity. In: MobileHealth '11: Proceedings of the First ACM MobiHoc Workshop on Pervasive Wireless Healthcare. 2011 Presented at: MobiHoc '11: The Twelfth ACM International Symposium on Mobile Ad Hoc Networking and Computing; May 16, 2011; Paris, France p. 1-5. [doi: 10.1145/2007036.2007048]
46. Hammersley ML, Jones RA, Okely AD. Time2bHealthy - An online childhood obesity prevention program for preschool-aged children: A randomised controlled trial protocol. Contemp Clin Trials 2017 Oct;61:73-80. [doi: 10.1016/j.cct.2017.07.022] [Medline: 28739536]
47. Napolitano MA, Whiteley JA, Mavredes MN, Faro J, DiPietro L, Hayman LL, et al. Using social media to deliver weight loss programming to young adults: Design and rationale for the Healthy Body Healthy U (HBHU) trial. Contemp Clin Trials 2017 Sep;60:1-13 [FREE Full text] [doi: 10.1016/j.cct.2017.06.007] [Medline: 28611007]
48. 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]
49. Tweepy Documentation. Tweepy. URL: https://docs.tweepy.org/en/stable/ [accessed 2022-07-13]
50. Turner-McGrievy GM, Beets MW. Tweet for health: using an online social network to examine temporal trends in weight loss-related posts. Transl Behav Med 2015 Jun;5(2):160-166 [FREE Full text] [doi: 10.1007/s13142-015-0308-1] [Medline: 26029278]
51. Danescu-Niculescu-Mizil C, West R, Jurafsky D, Leskovec J, Potts C. No country for old members: user lifecycle and linguistic change in online communities. In: WWW '13: Proceedings of the 22nd international conference on World Wide Web. 2013 Presented at: WWW '13: 22nd International World Wide Web Conference; May 13-17, 2013; Rio de Janeiro, Brazil p. 307-318. [doi: 10.1145/2488388.2488416]
52. Park A, Hartzler AL, Huh J, McDonald DW, Pratt W. Automatically Detecting Failures in Natural Language Processing Tools for Online Community Text. J Med Internet Res 2015 Aug 31;17(8):e212 [FREE Full text] [doi: 10.2196/jmir.4612] [Medline: 26323337]
53. So J, Prestin A, Lee L, Wang Y, Yen J, Chou WS. What Do People Like to "Share" About Obesity? A Content Analysis of Frequent Retweets About Obesity on Twitter. Health Commun 2016;31(2):193-206. [doi: 10.1080/10410236.2014.940675] [Medline: 26086083]
54. Guest G, MacQueen KM, Namey EE. Applied Thematic Analysis. New York, NY: SAGE Publications; 2012.
55. Strauss AL, Corbin JM. Basics of qualitative research. Newbury Park, CA: Sage Publications; 1990.
56. Bruckman A. Teaching Students to Study Online Communities Ethically. J Infor Ethics 2006 Nov 1;15(2):82-98 [FREE Full text]
57. Blei D, Ng A, Jordan M. Latent Dirichlet Allocation. J Mach Learn Res 2003:993-1022 [FREE Full text]
58. Park A, Conway M. Harnessing Reddit to Understand the Written-Communication Challenges Experienced by Individuals With Mental Health Disorders: Analysis of Texts From Mental Health Communities. J Med Internet Res 2018 Apr 10;20(4):e121 [FREE Full text] [doi: 10.2196/jmir.8219] [Medline: 29636316]
59. Bird S. NLTK: the natural language toolkit. In: COLING-ACL '06: Proceedings of the COLING/ACL on Interactive presentation sessions. 2006 Presented at: COLING/ACL 2006 Interactive Presentations; July 17-18, 2006; Sydney, Australia p. 69-72. [doi: 10.3115/1225403.1225421]
60. Rehurek R, Sojka P. Software Framework for Topic Modelling with Large Corpora. In: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks. 2010 Presented at: LREC 2010 Workshop on New Challenges for NLP Frameworks; May 22, 2010; Valletta, Malta p. 45-50.
61. Wordcloud: Project Description. Python Software Foundation. URL: https://pypi.python.org/pypi/wordcloud [accessed 2022-07-13]
62. Hartzler AL, Weis B, Cahill C, Pratt W, Park A, Backonja U, et al. Design and Usability of Interactive User Profiles for Online Health Communities. ACM Trans Comput Hum Interact 2016 Jul 21;23(3):1-33. [doi: 10.1145/2903718]
63. Chen H, Kelliher A. Conversational Lives: Visualizing Interpersonal Online Social Interactions. In: Online Communities and Social Computing. 2011 Presented at: OCSC 2011: Online Communities and Social Computing: 4th International Conference; Jul 9-14, 2011; Orlando, FL p. 241-250. [doi: 10.1007/978-3-642-21796-8_26]
64. Bastian M, Heymann S, Jacomy M. Gephi: An Open Source Software for Exploring and Manipulating Networks. In: Proceedings of the International AAAI Conference on Web and Social Media. 2009 Presented at: Third International AAAI Conference on Weblogs and Social Media; May 17-20, 2009; San Jose, CA p. 361-362. [doi: 10.1007/978-1-4614-6170-8_299]
65. Martin S, Brown WM, Klavans R, Boyack KW. OpenOrd: an open-source toolbox for large graph layout. Visualization and Data Analysis. URL: https://www.academia.edu/25474162/OpenOrd_An_open_source_toolbox_for_large_graph_layout [accessed 2022-07-14]

66. Gore RJ, Diallo S, Padilla J. You Are What You Tweet: Connecting the Geographic Variation in America's Obesity Rate to Twitter Content. PLoS One 2015;10(9):e0133505 [FREE Full text] [doi: 10.1371/journal.pone.0133505] [Medline: 26332588]
67. Fruchterman TMJ, Reingold EM. Graph drawing by force-directed placement. Softw Pract Exper 1991 Nov;21(11):1129-1164. [doi: 10.1002/spe.4380211102]
68. 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 Oct 16;16(10):e238 [FREE Full text] [doi: 10.2196/jmir.3622] [Medline: 25320863]
69. Ratkiewicz J, Conover M, Meiss M, Gonçalves B, Patil S, Flammini A, et al. Truthy: Mapping the Spread of Astroturf in Microblog Streams. In: Proceedings of the 20th international conference companion on World wide web. 2011 Presented at: WWW '11: 20th International World Wide Web Conference; Mar 28-Apr 1, 2011; Hyderabad, India p. 249-252. [doi: 10.1145/1963192.1963301]
70. Hartzler A, Pratt W. Managing the personal side of health: how patient expertise differs from the expertise of clinicians. J Med Internet Res 2011 Aug 16;13(3):e62 [FREE Full text] [doi: 10.2196/jmir.1728] [Medline: 21846635]
71. Oh S, Worrall A. Health answer quality evaluation by librarians, nurses, and users in social Q&A. Libr Inf Sci Res 2013 Oct;35(4):288-298. [doi: 10.1016/j.lisr.2013.04.007]
72. Newman MW, Lauterbach D, Munson SA, Resnick P, Morris ME. It's not that I don't have problems, I'm just not putting them on Facebook: Challenges and opportunities in using online social networks for health. In: CSCW 11 Proceedings of ACM 2011 Conference on Computer Supported Cooperative Work. 2011 Presented at: ACM 2011 Conference on Computer Supported Cooperative Work; Mar 19-23, 2011; Hangzhou, China p. 341-350. [doi: 10.1145/1958824.1958876]
73. Shpigelman C, Weiss PLT, Reiter S. E-Mentoring for All. Comput Human Behav 2009 Jul;25(4):919-928. [doi: 10.1016/j.chb.2009.03.007]
74. Hwang KO, Ning J, Trickey AW, Sciamanna CN. Website usage and weight loss in a free commercial online weight loss program: retrospective cohort study. J Med Internet Res 2013 Jan 15;15(1):e11 [FREE Full text] [doi: 10.2196/jmir.2195] [Medline: 23322819]
75. Millen D, Patterson J. Stimulating social engagement in a community network. In: Proceedings of the 2002 ACM conference on Computer supported cooperative work. 2002 Presented at: CSCW02: Computer Supported Cooperative Work; Nov 16-20, 2002; New Orleans, LA p. 306-313. [doi: 10.1145/587078.587121]
76. Joyce E, Kraut RE. Predicting Continued Participation in Newsgroups. J Comp Mediated Comm 2006 Apr;11(3):723-747. [doi: 10.1111/j.1083-6101.2006.00033.x]
77. Wang Y, Kraut R, Levine J. To stay or leave?: the relationship of emotional and informational support to commitment in online health support groups. In: Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work. 2012 Presented at: CSCW '12: Computer Supported Cooperative Work; Feb 11-15, 2012; Seattle, WA p. 833-842. [doi: 10.1145/2145204.2145329]
78. Arguello J, Butler B, Joyce L, Kraut R, Ling K, Wang X. Talk to me: foundations for successful individual-group interactions in online communities. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2006 Presented at: CHI06: CHI 2006 Conference on Human Factors in Computing Systems; Apr 22-27, 2006; Montréal, QC p. 959-968. [doi: 10.1145/1124772.1124916]
79. Nonnecke B, Preece J. Lurker demographics: counting the silent. In: Proceedings of the SIGCHI conference on Human Factors in Computing Systems. 2000 Presented at: CHI00: Human Factors in Computing Systems; Apr 1-6, 2000; The Hague, Netherlands p. 73-80. [doi: 10.1145/332040.332409]
80. Park A, Hartzler AL, Huh J, McDonald DW, Pratt W. Homophily of Vocabulary Usage: Beneficial Effects of Vocabulary Similarity on Online Health Communities Participation. AMIA Annu Symp Proc 2015;2015:1024-1033 [FREE Full text] [Medline: 26958240]
81. Vayena E, Salathé M, Madoff LC, Brownstein JS. Ethical challenges of big data in public health. PLoS Comput Biol 2015 Feb;11(2):e1003904 [FREE Full text] [doi: 10.1371/journal.pcbi.1003904] [Medline: 25664461]
82. Mikal J, Hurst S, Conway M. Ethical issues in using Twitter for population-level depression monitoring: a qualitative study. BMC Med Ethics 2016 Apr 14;17:22 [FREE Full text] [doi: 10.1186/s12910-016-0105-5] [Medline: 27080238]
83. Padilla JJ, Kavak H, Lynch CJ, Gore RJ, Diallo SY. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. PLoS One 2018;13(6):e0198857 [FREE Full text] [doi: 10.1371/journal.pone.0198857] [Medline: 29902270]
84. Smith A, Anderson M. Social Media Use in 2018. Pew Research Center. URL: http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/; [accessed 2022-07-10]

Abbreviations

LDA: latent Dirichlet allocation.

Edited by A Mavragani; submitted 15.06.19; peer-reviewed by MM Amini, ORCID: 0000-0002-5279-213X, J Mikal, R Gore, E Fernandes de Mello Araujo, J Wright; comments to author 08.08.20; revised version received 04.01.21; accepted 23.06.22; published 20.07.22.

Please cite as:
Park A
Tweets Related to Motivation and Physical Activity for Obesity-Related Behavior Change: Descriptive Analysis
J Med Internet Res 2022;24(7):e15055
URL: https://www.jmir.org/2022/7/e15055
doi:10.2196/15055
PMID:35857347

Original Paper

Pandemic-Triggered Adoption of Telehealth in Underserved Communities: Descriptive Study of Pre- and Postshutdown Trends

Pei Xu[1], PhD; Matthew Hudnall[2], PhD; Sidi Zhao[1]; Uzma Raja[1], PhD; Jason Parton[2], PhD; Dwight Lewis[2], PhD

[1]Department of Systems and Technology, Auburn University, Auburn, AL, United States

[2]Department of Information Systems, Statistics, and Management Science, University of Alabama, Tuscaloosa, AL, United States

Corresponding Author:
Matthew Hudnall, PhD
Department of Information Systems, Statistics, and Management Science
University of Alabama
500 Colonial Drive
Tuscaloosa, AL, 35487
United States
Phone: 1 205 348 0856
Email: matthew.hudnall@ua.edu

Abstract

Background: The adoption of telehealth services has been a challenge in rural communities. The reasons for the slow adoption of such technology-driven services have been attributed to social norms, health care policies, and a lack of infrastructure to support the delivery of services. However, the COVID-19 pandemic–related shutdown of in-person health care services resulted in the usage of telehealth services as a necessity rather than a choice. The pandemic also fast-tracked some needed legislation to allow medical cost reimbursement for remote examination and health care services. As services return to normalcy, it is important to examine whether the usage of telehealth services during the period of a shutdown has changed any of the trends in the acceptance of telehealth as a reliable alternative to traditional in-person health care services.

Objective: Our aim was to explore whether the temporary shift to telehealth services has changed the attitudes toward the usage of technology-enabled health services in rural communities.

Methods: We examined the Medicaid reimbursement data for the state of Alabama from March 2019 through June 2021. Selecting the telehealth service codes, we explored the adoption rates in 3 phases of the COVID-19 shutdown: prepandemic, pandemic before the rollout of mass vaccination, and pandemic after the rollout of mass vaccination.

Results: The trend in telemedicine claims had an opposite pattern to that in nontelemedicine claims across the 3 periods. The distribution of various characteristics of patients who used telemedicine (age group, gender, race, level of rurality, and service provider type) was different across the 3 periods. Claims related to behavior and mental health had the highest rates of telemedicine usage after the onset of the pandemic. The rate of telemedicine usage remained at a high level after the rollout of mass vaccination.

Conclusions: The current trends indicate that adoption of telehealth services is likely to increase postpandemic and that the consumers (patients), service providers, health care establishments, insurance companies, and state and local policies have changed their attitudes toward telehealth. An increase in the use of telehealth could help local and federal governments address the shortage of health care facilities and service providers in underserved communities, and patients can get the much-needed care in a timely and effective manner.

(J Med Internet Res 2022;24(7):e38602) doi:10.2196/38602

KEYWORDS

telehealth services; telemedicine; COVID-19; rural communities; mental health; Medicaid; health service; telehealth; health care service; health care facility; undeserved community; undeserved population; health claim; technology adoption; health insurance

Introduction

The COVID-19 pandemic has stress-tested worldwide health systems like no other event in modern times. Requirements for limited travel, social distancing, and business closures have adversely impacted many segments of society in order to help mitigate the spread of the virus. In this context, telemedicine

has been promoted and expanded to reduce the risk of viral transmission [1].

Telemedicine enhances the delivery and availability of health care services. As a result, during the COVID-19 pandemic, telemedicine should be a vital tool in providing care, while keeping patients and health professionals safe [2]. Telemedicine, as an effective tool to increase patients' accessibility to health services, may effectively reduce medical costs and improve patients' quality of life [3]. The future benefits of telemedicine include its cost-effectiveness, its ability to expand specialty services, and its potential to help alleviate looming physician shortages [4].

Prior to COVID-19, telehealth in Alabama, USA, was broadly available but seldom utilized. Telehealth equipment was available in all 67 county health offices, most hospitals, and many physicians throughout the state. Within the Medicaid population, prepandemic telehealth utilization averaged around 1000 people, with 1.3 sessions per person within any given month [5]. This was among a Medicaid population that consists of approximately 1.2 million persons on an annual basis. This low utilization rate was due to a myriad of factors related to billing practice, patient/physician comfort, and familiarity with using the technology [6,7].

On April 3, 2020, Alabama Governor Kay Ivey issued a stay-at-home order to help curb the increasing spread of the virus [8]. This order, in conjunction with the widespread fear of contracting the virus during the prior months, created a unique environment for the necessitated rapid adoption of telemedicine. At the same time, insurance companies adjusted their policies so that telemedicine visits became reimbursable at approximately the same rates as in-person visits [9]. The changes to reimbursement rules coupled with the necessitated modality created the environment through which telemedicine was given an opportunity to shine.

Prior studies have highlighted the importance of studying the granularity of telemedicine adoption using subpopulation analyses. For instance, Chu et al [10] suggested that further studies are required "to assess the potential barriers to telemedicine experienced by rural populations compared to those experienced by urban populations and the impact of telemedicine compared to that of in-person care on other forms of health care utilization, outcomes, and quality of care among vulnerable and at-risk patient groups in the rural population." Monaghesh and Hajizadeh [2] proposed that researchers can "examine the effectiveness of using telehealth approaches in different health areas, especially in the field of home nursing the elderly who are high-risk people in the community. It is also highly recommended to use this technology in the field of psychiatry as it does not require in-person visits." As a response to these calls, this paper aims to examine the prior utilization rates of telemedicine among the Alabama Medicaid population and compares those rates stratified by demographics to the rates during the COVID-19 pandemic.

Methods

Data

Since 2014, the Institute of Data and Analytics (IDA) at the University of Alabama has had a consulting contract with Alabama Medicaid, in which the IDA supports Alabama Medicaid's analytics group. To fulfill this contract, the IDA has maintained a secure copy of all Alabama Medicaid enrollment and claims data since 2010. These data include specific billing/claim information for all enrolled in Alabama's Medicaid program. Annually, approximately 1 (25%) of every 4 Alabamians has some form of Medicaid coverage during the year. Medicaid insurance services are made available to low-income populations, as well as those who are disabled, at no cost to participants.

Data from the procedure and provider tables within the Medicaid database were extracted that contained corresponding billing codes indicating actions had taken place over telemedicine. Claims information within the Medicaid database can lag behind for up to 6 months since facilities and practitioners may not always bill in a timely manner. However, over 80% of claims are received within a month from the time of service. The variables extracted included the year and month of the service, the county where the service took place (provider location), the gender of the patient, the age range of the patient corresponding to US Census age groups, provider/procedure information, modality of service (in person vs telehealth), and total billing costs paid. Procedures were further grouped by type of procedure and group of procedure, and providers were further grouped by type of provider as well as the specialty of the provider. These variables were aggregated, and calculations of counts of unique persons, counts of claims, and total cost of claims for each level of aggregation were calculated. The data represent a complete sampling of the Alabama Medicaid records from March 1, 2019, through June 30, 2021. The data were extracted from the August 2021 data backup of the Alabama Medicaid data, which included transactions through July 2021 and represented approximately 80% of all claims through June of that year. In total, 94,254,599 patient claims covered by Alabama Medicaid at some point during the analysis period are represented in the telehealth utilization data, along with a total of about US $5 billion across 738 procedure types from 27 provider categories across 97 specialties.

The COVID-19 tracking reports were downloaded from the COVID Tracking Project [11], which collects and publishes the most comprehensive data about COVID-19 in the United States. We also collected the COVID-19 vaccination progress report for Alabama from the USAFacts website [12]. Lastly, a classification of areas as being rural or urban at the county level was obtained from the Alabama Rural Health Association [13].

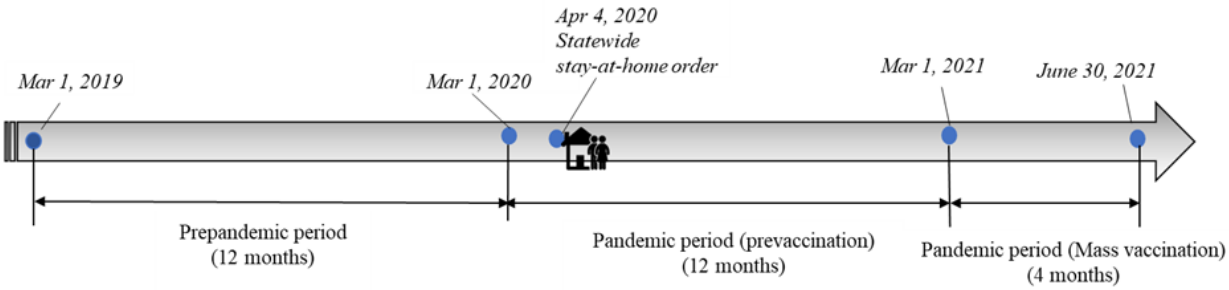
Analysis

The cumulative time frame of analysis was from March 1, 2019, through June 30, 2021, as shown in Figure 1. Alabama announced the first known cases of coronavirus in March 2020; therefore, the time frame being considered as pre-COVID-19 was the period from March 1, 2019, to February 29, 2020. Broad COVID-19 vaccine availability in Alabama occurred during

March 2021, so the time frame from March 1, 2021, to June 30, 2021 was analyzed separately as a mass vaccination condition. The core pandemic time frame was characterized by March 1, 2020, through February 28, 2021. This allowed for 12 months of data prior to COVID-19, 12 months of data during the pandemic, and 4 months of analysis of the time frame when vaccines were widely available.

We compared the characteristics of patients who utilized and did not utilize the telemedicine service in each period. For each telemedicine usage modality, chi-square tests were conducted to assess the distribution of services and patient characteristics (eg, age group, gender, race, level of rurality, and service provider type) across different study periods. We then calculated the rate of telemedicine usage for the various characteristics for each month in the study periods.

Figure 1. Illustration of study periods.



Ethical Considerations

Institutional board approval was sought and obtained for this project from the University of Alabama's ethics board (UA IRB protocol #21-05-4661). The full study protocol is available through the institution or by contacting the authors. Additionally, the research was vetted and approved by Alabama Medicaid's internal research review process.

Results

Overview

The data set contained 94,254,599 Medicaid claims from March 1, 2019, through June 30, 2021, in the state of Alabama. For each period, we reported the rate of telemedicine usage, which increased from 0.12% (n=55,613) in the prepandemic period to 3.43% (n=1,141,282) and 1.85% (n=254,807) in the subsequent 2 periods, respectively, as seen in Table 1. In addition, the second period (ie, the first year after the onset of the pandemic) saw the lowest average monthly Medicaid claims (2,774,171 claims per month) but the highest average monthly telemedicine Medicaid claims (95,107 claims per month) among the 3 study periods. Such an opposite trend was partly due to the statewide stay-at-home orders in that period.

Figure 2 provides an overview of monthly changes in new COVID-19 confirmed cases, vaccination progress, and telemedicine usage in Alabama. As can be seen in the figure, the nontelemedicine claims and total claims dramatically dropped since March 2020. Both lines reached the valley when the state issued a stay-at-home order in April 2020, while at the same time, telemedicine usage reached its peak point. The monthly claims of telemedicine usage gradually decreased during the months after April 2020, but the number of monthly claims was still at a relatively high level at the end of our study period (ie, June 2021). This is an indication that telemedicine usage is still popular, even after the rollout of mass vaccination and the relaxation of the COVID-19–related executive orders. Interestingly, the monthly new confirmed cases peaked in December 2020, but neither telemedicine nor nontelemedicine claims exhibited any obvious changes during that period. This is possibly due to people having become accustomed to the new norm brought by the pandemic.

In Table 2 (Multimedia Appendix 1) and Table 3 (Multimedia Appendix 2), we summarize the characteristics of Medicaid claims in the 3 periods (ie, before the pandemic, after the onset of the pandemic, and after the rollout of mass vaccination). Monthly Medicaid claims are reported along with the proportions for the groups within each classification. The sparklines, with the lowest and highest values marked in blue dots, were used to visualize the trend in claim counts over the 3 periods. Chi-square tests were performed to assess the distribution of various characteristics of patients who used telemedicine (age group, gender, race, level of rurality, and service provider type) across the 3 periods. These chi-square tests all resulted in a P value <.001. Applying the correction for multiple comparisons, we calculated the Bonferroni correction α values for each test, and the smallest adjusted α value was 0.05/42=0.0012, suggesting that there were significant shifts in the constitution of telemedicine and nontelemedicine usage in terms of the aforementioned patient/service characteristics classifications. The trend in the total number of Medicaid claims was similar to nontelemedicine Medicaid claims, as the latter made up a large proportion of the former (see Table 1).

Table 1. Medicaid claims in each study period.

COVID-19 periods	Time frames	Total Medicaid claims			Average monthly Medicaid claims		
		Telemedicine, n (%)	Nontelemedicine, n (%)	Total, N	Telemedicine, n (%)	Nontelemedicine, n (%)	Total, N
Prepandemic	March 1, 2019-February 29, 2020	55,613 (0.12)	47,110,415 (99.88)	47,166,028	4634 (0.11)	3,925,868 (99.89)	3,930,502
Pandemic (prevaccination rollout)	March 1, 2020-February 28, 2021	1,141,282 (3.43)	32,148,768 (96.57)	33,290,050	95,107 (3.43)	2,679,064 (96.57)	2,774,171
Pandemic (postvaccination rollout)	March 1, 2021-June 30, 2021	254,807 (1.85)	13,543,714 (98.15)	13,798,521	63,702 (1.85)	3,385,929 (98.15)	3,449,630

Figure 2. Overview of monthly COVID-19 cases, vaccination, and telemedicine usage in Alabama. The "Total Claims" and "Nontelemedicine Claims" series are plotted on the right-hand axis, while the other data series are plotted on the left-hand axis.

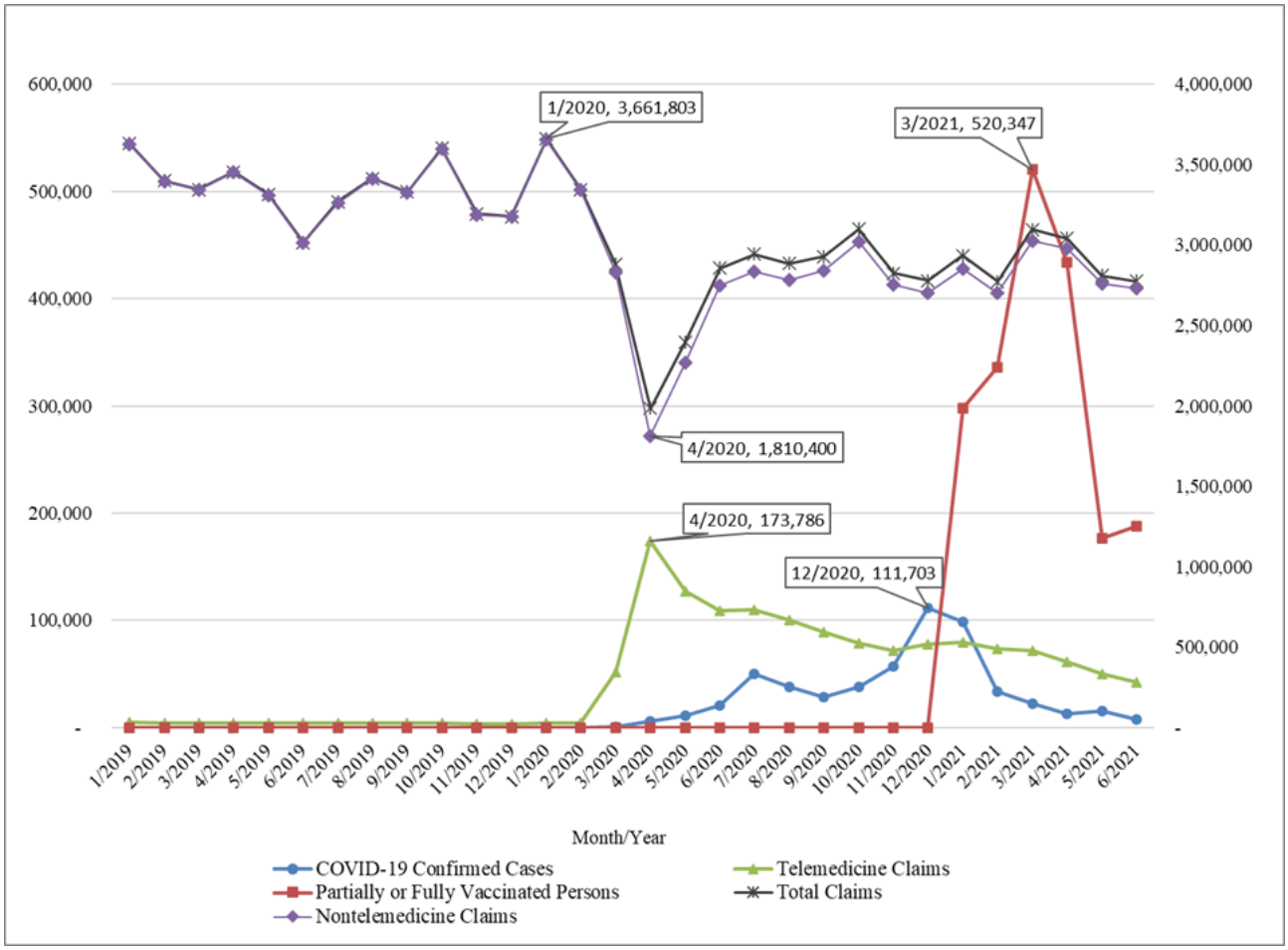


Table 2. Monthly telemedicine and nontelemedicine Medicaid claims grouped by period and patient demographics.

Variables	Monthly nontelemedicine Medicaid claims, n (%)			Monthly telemedicine Medicaid claims, n (%)		
	Period 1	Period 2	Period 3	Period 1	Period 2	Period 3
Age (years), *P*<.001						
0-17	1,621,283 (41.30)	1,073,173 (40.06)	1,500,784 (44.32)	1733 (37.39)	51,071 (53.70)	35,601 (55.89)
18-29	462,397 (11.78)	332,697 (12.42)	444,775 (13.14)	457 (9.87)	9868 (10.38)	6693 (10.51)
30-39	345,943 (8.81)	249,813 (9.32)	312,790 (9.24)	557 (12.01)	9929 (10.44)	6617 (10.39)
40-49	316,603 (8.06)	219,858 (8.21)	268,471 (7.93)	578 (12.48)	7892 (8.30)	5153 (8.09)
50-64	680,963 (17.35)	460,029 (17.17)	525,477 (15.52)	1000 (21.57)	12,617 (13.27)	7795 (12.24)
65-74	284,218 (7.24)	197,950 (7.39)	195,061 (5.76)	236 (5.09)	2501 (2.63)	1338 (2.10)
75-84	141,137 (3.60)	95,784 (3.58)	92,449 (2.73)	59 (1.28)	842 (0.88)	376 (0.59)
85+	73,325 (1.87)	49,759 (1.86)	46,122 (1.36)	14 (0.31)	388 (0.41)	129 (0.20)
Gender, *P*<.001						
Female	2,399,426 (61.12)	1,649,587 (61.57)	2,088,520 (61.68)	2400 (51.79)	53,068 (55.80)	35,260 (55.35)
Male	1,517,497 (38.65)	1,020,126 (38.08)	1,282,413 (37.87)	2223 (47.98)	41,945 (44.10)	28,377 (44.55)
Unknown	8945 (0.23)	9351 (0.35)	14,996 (0.44)	11 (0.24)	94 (0.10)	65 (0.10)
Race/ethnicity, *P*<.001						
Black	1,472,153 (37.50)	1,009,310 (37.67)	1,244,880 (36.77)	1808 (39.01)	33,234 (34.94)	23,437 (36.79)
Hispanic	114,738 (2.92)	83,240 (3.11)	121,757 (3.60)	49 (1.06)	2100 (2.21)	1317 (2.07)
Other	671,783 (17.11)	466,643 (17.42)	608,547 (17.97)	796 (17.17)	18,772 (19.74)	12,792 (20.08)
White	1,666,694 (42.45)	1,119,461 (41.79)	1,409,934 (41.64)	1981 (42.75)	40,991 (43.10)	26,152 (41.05)
Rurality, *P*<.001						
High	1,236,296 (31.49)	828,028 (30.91)	1,025,477 (30.29)	2145 (46.28)	27,939 (29.38)	18,370 (28.84)
Moderate	713,244 (18.17)	489,683 (18.28)	631,993 (18.67)	1025 (22.11)	17,849 (18.77)	11,990 (18.82)
Low	1,974,916 (50.31)	1,360,294 (50.77)	1,726,747 (51.00)	1464 (31.58)	49,288 (51.82)	33,308 (51.29)

Table 3. Monthly telemedicine and nontelemedicine Medicaid claims grouped by period and provider type.

Provider type	Monthly nontelemedicine Medicaid claims (P<.001), n (%)			Monthly telemedicine Medicaid claims (P<.001), n (%)		
	Period 1	Period 2	Period 3	Period 1	Period 2	Period 3
American Academy of Physician Associates (AAPA)–employed physicians	23,264 (0.74)	15,387 (0.73)	21,853 (0.79)	0	355 (0.41)	300 (0.50)
Behavioral health	11,081 (0.35)	8000 (0.38)	14,764 (0.54)	18 (0.43)	4883 (5.59)	4299 (7.15)
Case manager (targeted)	33,603 (1.07)	22,280 (1.06)	23,112 (0.84)	10 (0.24)	514 (0.59)	327 (0.54)
Certified registered nurse anesthetist (CRNA)/certified registered nurse practitioner (CRNP)/nurse/midwife	271,499 (8.67)	189,631 (9.04)	262,797 (9.53)	206 (4.95)	6072 (6.95)	4374 (7.28)
Dentist	230,189 (7.35)	182,720 (8.71)	288,609 (10.47)	0	25 (0.03)	4 (0.01)
Federally qualified health clinic (FQHC)	141,790 (4.53)	93,818 (4.47)	126,486 (4.59)	83 (2.00)	4404 (5.04)	2224 (3.7)
Hospital	170,595 (5.45)	105,565 (5.03)	103,996 (3.77)	1 (0.01)	18 (0.02)	4 (0.01)
Mental health	262,456 (8.39)	155,685 (7.42)	185,949 (6.75)	3010 (72.33)	34,648 (39.64)	26,415 (43.96)
Optometrist	75,384 (2.41)	50,065 (2.39)	70,023 (2.54)	0	22 (0.02)	2 (0.00)
Physician	1,681,678 (53.73)	1,117,353 (53.29)	1,438,575 (52.18)	821 (19.73)	23,245 (26.6)	13,408 (22.31)
Podiatrist	6788 (0.22)	3152 (0.15)	3100 (0.11)	0	1 (0.001)	1 (0.001)
Psychologist	17,821 (0.57)	5551 (0.26)	8173 (0.30)	10 (0.23)	4659 (5.33)	3996 (6.65)
Rural health clinic	132,971 (4.25)	91,530 (4.37)	126,822 (4.60)	2 (0.04)	5047 (5.77)	3099 (5.16)
Therapist	70,612 (2.26)	56,159 (2.68)	82,554 (2.99)	1 (0.02)	3503 (4.01)	1645 (2.74)

Age

We followed the age group classification used by the Centers for Disease Control and Prevention (CDC) in reporting COVID-19 cases. We observed a significant use of telemedicine in all age groups after the onset of the pandemic. Prior to the pandemic, a greater proportion of minors (aged 0-17 years) used telemedicine (1733/4634, 37.39%) compared to other age groups, and this proportion was even more significant after the onset of the pandemic (period 2: 51,071/95,107, 53.70%; period 3: 35,601/21,234, 55.89%). For both telemedicine and nontelemedicine services, the monthly claims were the highest in the first period and lowest in the second period for all groups except for adults aged older than 65 years.

Gender

Prior to the pandemic, the Medicaid claims of female and male groups were almost the same (2400, 51.79%, per month vs 2223, 47.98%, per month); however, after the onset of the pandemic and mass vaccination, the number of female telemedicine visits increased at a higher rate than male visits (period 2: 53,068, 55.80%, vs 41,945, 44.10%; period 3: 35,260, 55.35%, vs 28,377, 44.55%).

Race/Ethnicity

We observed a significant increase in the rate of telemedicine visits among all race groups in periods 2 and 3 compared to period 1. Such growth was least substantial in the African American community. The number of African American visits changed from 1808 (39.01%) to 33,234 (34.94%) from period 1 to period 2, and the latter was about 18 times greater than the former. The number of claims of the Hispanic group changed from 49 (1.06%) to 2100 (2.21%) per month from period 1 to period 2, which is a 42-fold increase. The White community also saw a 20-fold increase in terms of monthly claims.

Rurality

We first categorized the counties into 3 levels of rurality according to the method developed and used by the Alabama Rural Health Association [13]. The method uses 4 variables (ie, the percentage of total employment, the dollar value of agricultural production per square mile of land, the population per square mile of land, and the population of the largest city in the county), with each variable accounting for 25 of a possible 100 points. The higher the overall score is, the more rural a county is rated (Table 4).

For patients in all types of rurality, we observed an increase in telemedicine usage after the onset of the pandemic. Interestingly, the increase was more obvious in urban patients, as the proportion of urban telemedicine–using patients changed from 31.58% (n=1464) in period 1 to 51.82% (n=49,288) in period 2 and 52.29% (n=33,308) in period 3, while the overall urban telemedicine–using patients were around 50% in all 3 periods.

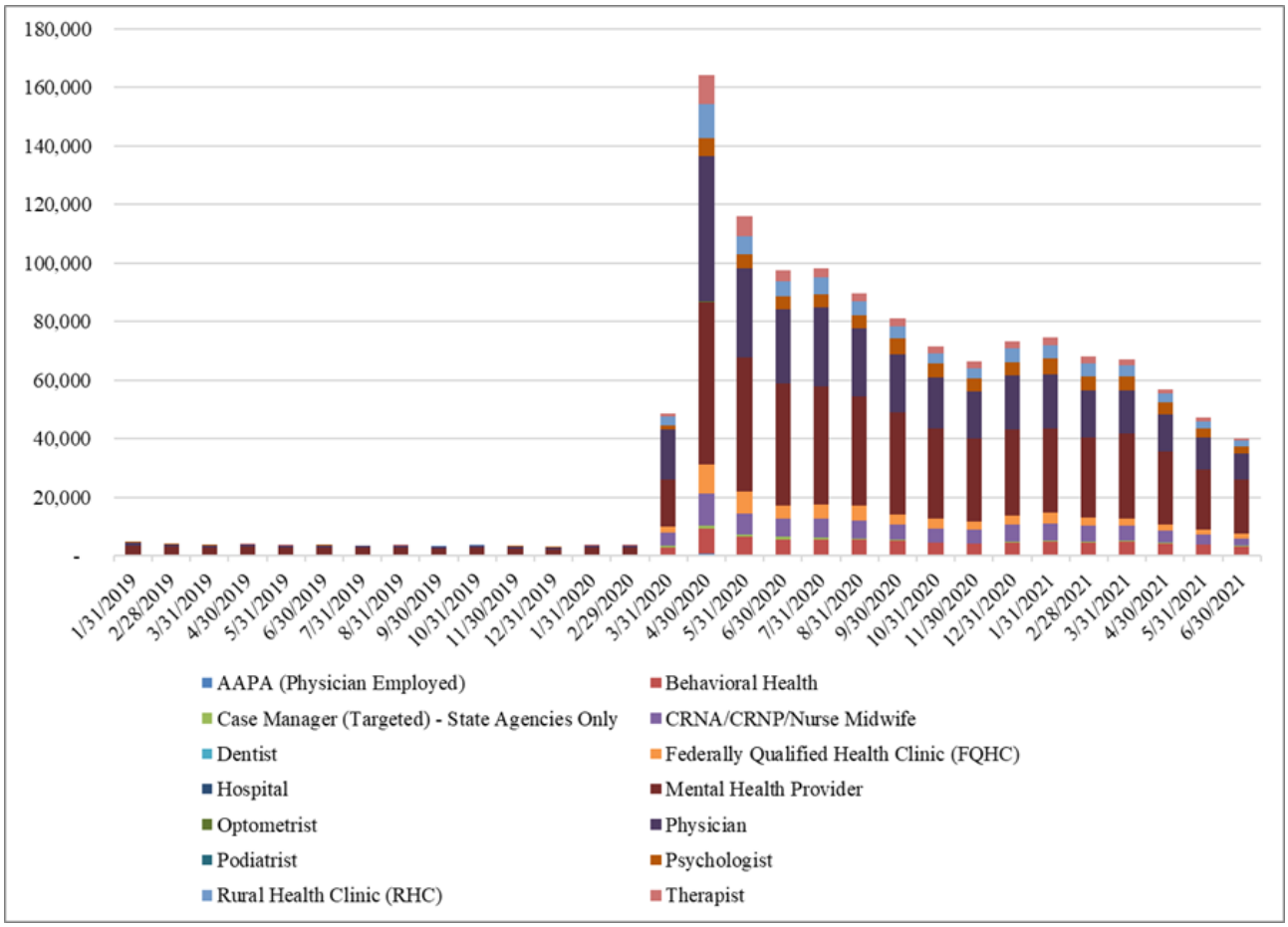
Table 4. Alabama rural and urban counties.

Level of rurality	Counties
Highly rural	Barbour, Bibb, Blount, Bullock, Butler, Cherokee, Choctaw, Clarke, Clay, Cleburne, Coffee, Conecuh, Coosa, Covington, Crenshaw, Cullman, Dallas, DeKalb, Escambia, Fayette, Franklin, Geneva, Greene, Hale, Henry, Jackson, Lamar, Lawrence, Lowndes, Macon, Marengo, Marion, Marshall, Monroe, Perry, Pickens, Pike, Randolph, Sumter, Washington, Wilcox, Winston
Moderately rural	Autauga, Baldwin, Chambers, Chilton, Colbert, Dale, Elmore, Limestone, Russell, St. Clair, Talladega, Tallapoosa, Walker
Urban	Calhoun, Etowah, Houston, Jefferson, Lauderdale, Lee, Madison, Mobile, Montgomery, Morgan, Shelby, Tuscaloosa

Provider Type

Table 3 (Multimedia Appendix 2) displays the monthly Medicaid claims by service type over the 3 study periods. The sparkline, with the lowest and highest values marked, indicates that the number of telemedicine claims for all service types dramatically increased in period 2 and slightly dropped in period 3. The number of a few services only slightly reduced in period 3, such as behavioral health, mental health, and psychology, suggesting a continued enthusiasm for telemedicine in these areas. In addition, mental health services accounted for about 72.33% (n=3010) of all telemedicine claims in period 1, 39.64% (n=34,648) in period 2, and 43.96% (n=26,415) in period 3. The shrinking in the mental health service proportion reveals the dramatic growth of telemedicine usage in other areas after the onset of the pandemic. Such a trend is also demonstrated in Figure 3, in which we visualize the distribution of claims across different service types.

Figure 3. Telemedicine claims by provider type.



Rate of Telemedicine Visits

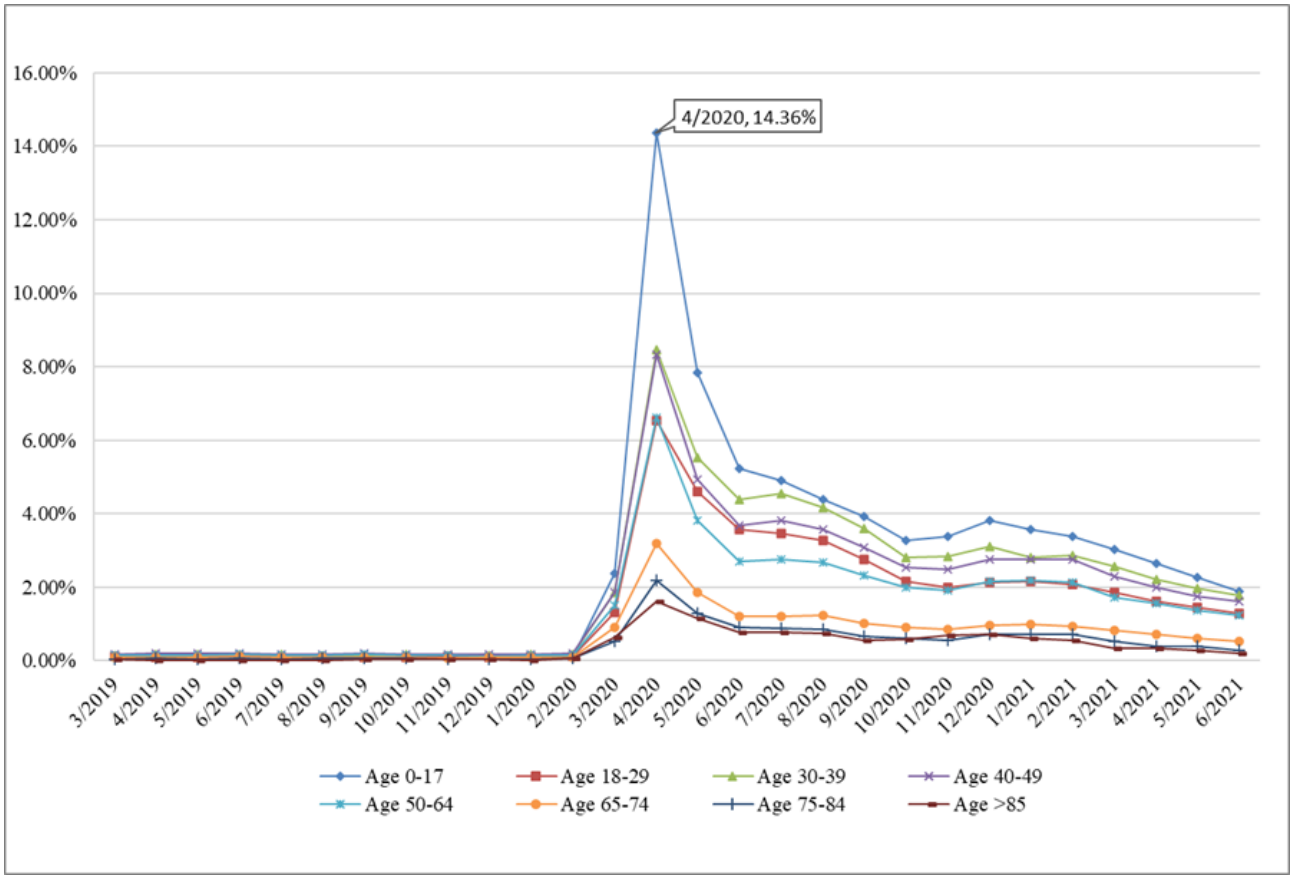
For each month within the study time frame, we reported the rate of telemedicine visits, which is defined as the count of telemedicine claims divided by the count of all claims. We then compared the rates across groups for the various characteristics.

Age

As shown in Figure 4, the rates of telemedicine visits were low (close to 0) for all age groups before the onset of the pandemic. We observed a significant increase in the rate of telemedicine visits among patients in all age groups after March 2020, especially the minor group (rate=14.36% in April 2020). In addition, the higher the age group, the lower the rate of telemedicine service. For instance, even in April 2020, the

telemedicine visit rate was only 1.61% for patients aged 85 years and older.

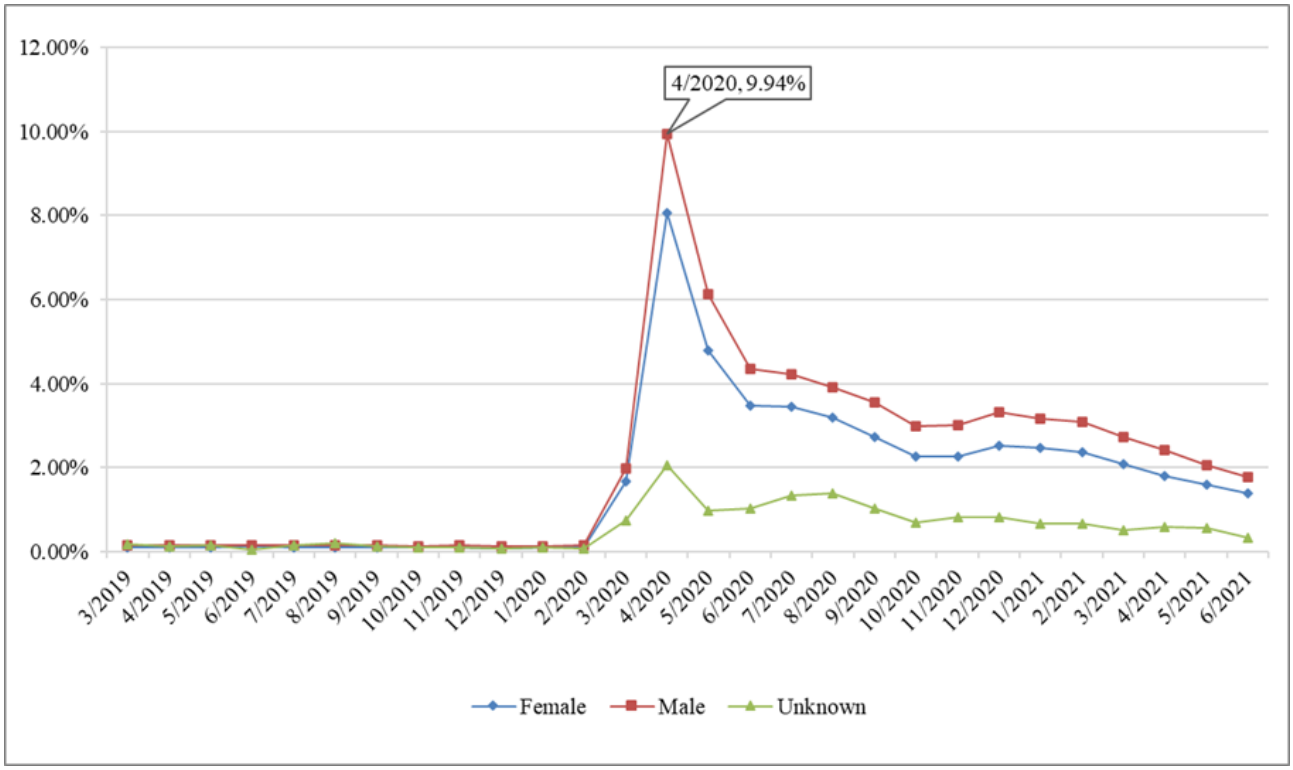
Figure 4. Rate of telemedicine visits by age group.



Gender

The rate of telemedicine visits increased significantly for both female and male groups, with the rate for the latter being slightly higher than for the former during the pandemic. This is likely because many female visits, such as labor and delivery, could not be conducted via telemedicine services. Moreover, males tend to have higher self-efficacy and trust in telemedicine technology compared to females and thus were more willing to switch to telemedicine services during the pandemic [14]. The trends are shown in Figure 5.

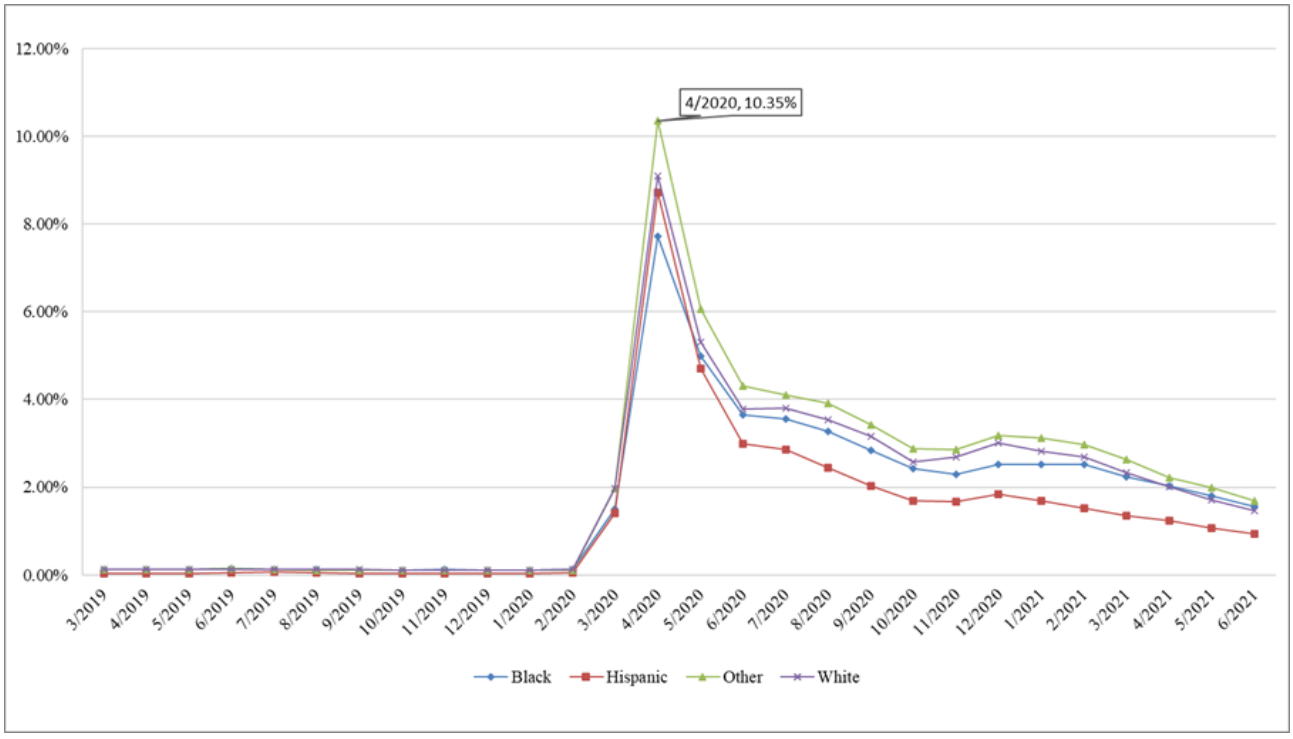
Figure 5. Rate of telemedicine visits by gender.



Race/Ethnicity

The rate of telemedicine visits (peak value=10.35%) was the highest for patients in race groups other than White, Black, and Hispanic. The Alabama Medicaid data system does not contain a field for ethnicity, so Hispanics are mapped as a race. The Hispanic group had the lowest telemedicine rate throughout the study periods except in April 2020, as shown in Figure 6.

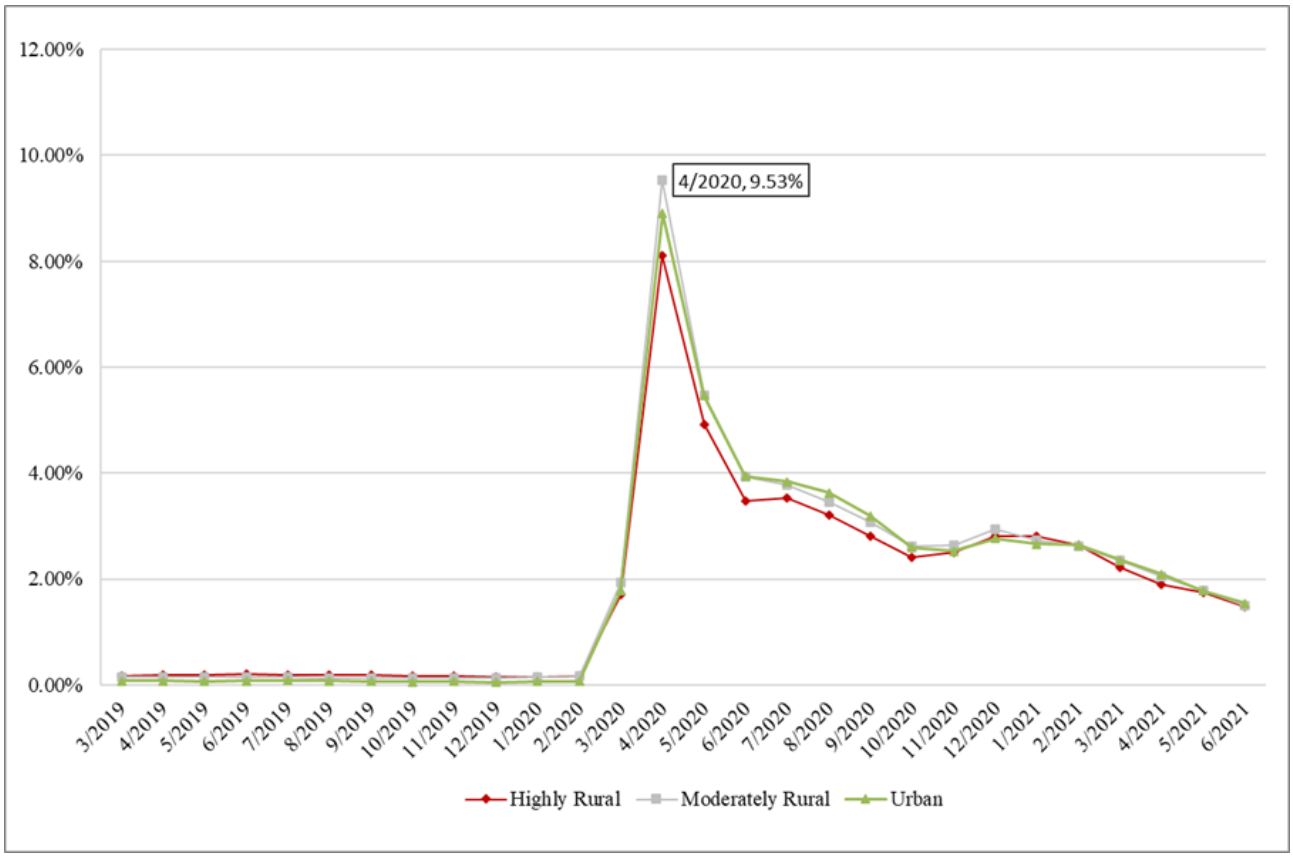
Figure 6. Rate of telemedicine visits by race.



Rurality

Like any other classifications, the rates of telemedicine visits were low for all rural levels, as presented in Figure 7. The onset of the pandemic has given rise to telemedicine visits across counties with different rurality levels. The trends in rates were similar for the 3 groups. The highest rate was observed for the moderately rural group in April 2020 (rate=9.53%).

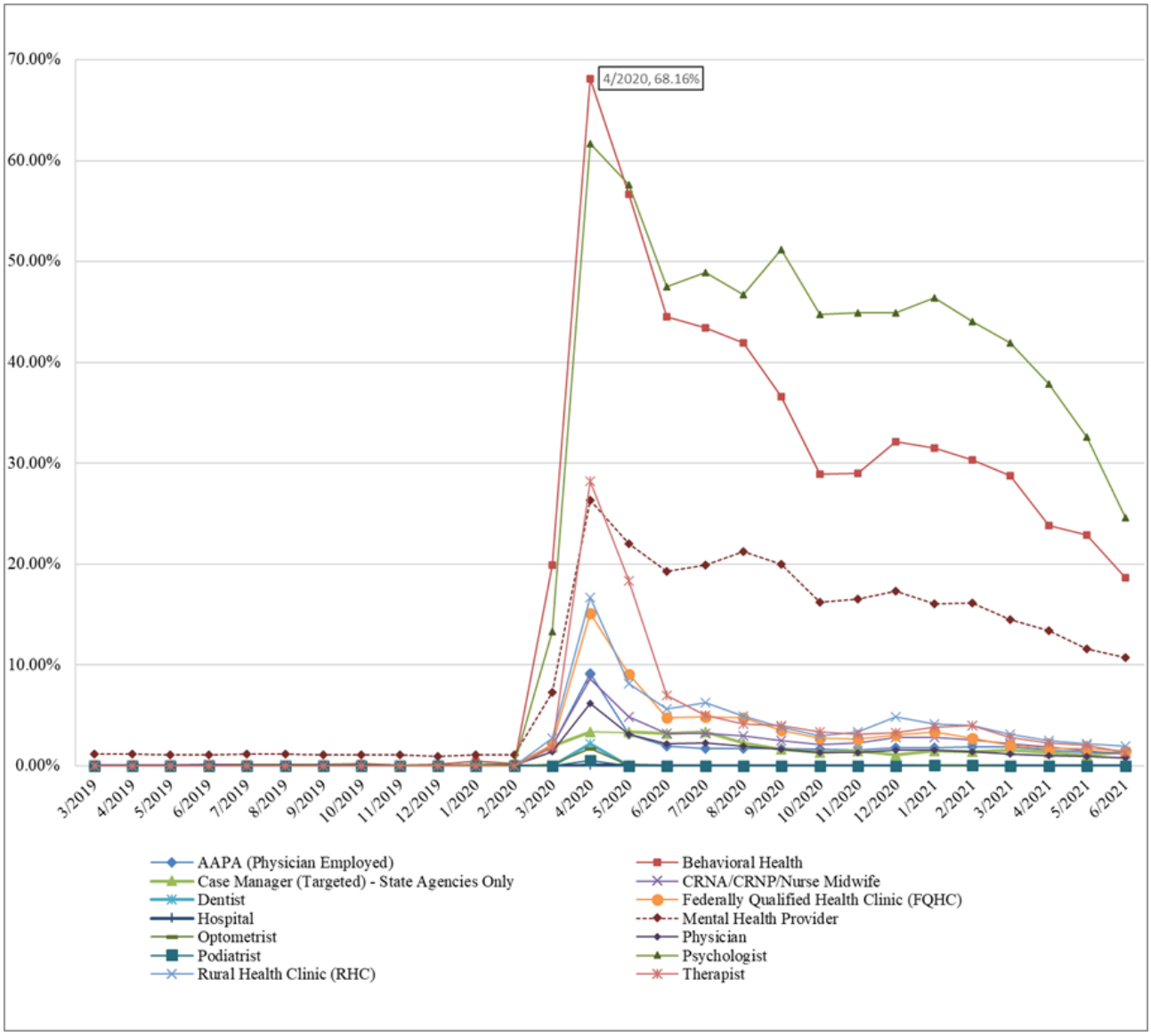
Figure 7. Rate of telemedicine visits by rurality.



Procedure Type

Although the rates of telemedicine visits for behavioral health providers, psychologists, and mental health providers were not high before the onset of the pandemic, these rates have surged since the onset of the pandemic. As seen from the peak values in April 2020, 68.16% (8419/12,352) of the visits for behavioral health–related conditions, 61.67% (5850/9486) of the services provided by a psychologist, and 26.35% (55,408/210,259) of the visits for mental health–related conditions were carried out via telemedicine. These rates were maintained at a relatively high value, even after the rollout of mass vaccination. In addition, we observed a sharp increase in telemedicine rates for other services right after the onset of the pandemic; however, these rates have dropped to below 5% since August 2020. The trends are shown in Figure 8.

Figure 8. Rate of telemedicine visits by provider type.



Discussion

Principal Findings

The COVID-19 pandemic introduced never-before-seen challenges to modern society. Although not directly planning for a pandemic, telehealth availability had been steadily increasing prepandemic but utilization was low. Barriers to implementing telemedicine include poor access to technical equipment, management, and reimbursement [15]. The top barriers are technology specific and could be overcome through training, change management techniques, and alternating delivery by telemedicine and personal patient-to-provider interaction. However, technology, as the major barrier, can be overcome through training, changing management techniques, and alternating delivery of services through telemedicine and personal patient-physician interactions [16]. The pandemic presented an opportunity to stress-test the underutilized telehealth modality, and the results showed that not only did the technology meet a large proportion of the needs but also a shift in society and providers toward a willingness to continue using the medium was observed. Beyond the baseline period, the COVID-19 pandemic has prompted a significant rise in the use of telemedicine for both urgent and nonurgent medical visits [17].

Although Wood et al [18] reported no significant differences in telehealth completion rates by age, sex, gender, or insurance, we found some differences in different dimensions. A telemedicine usage percentage of 57% in period 2 for 0-17-year-olds indicates a strong willingness of younger individuals to use telehealth services. The use of telemedicine by children varied by age, race, ethnicity, and recent preventive treatment, building upon previous concerns about disparities in telemedicine availability [19]. The extent to which this population directly uses telehealth versus having a parent or guardian facilitate the communications is unknown, and an area where expanded research could provide additional insights.

As a population, females experienced a higher increase in telehealth utilization compared to males. One possible reason for this could be related to the 0-17-year-old population. If mothers use telehealth for their children, then it stands to reason that they might likely do the same for themselves. Alabama

Medicaid covers more than half of all births annually in Alabama, so the overall population contains a large number of females and younger persons. According to Patel et al [20], rapid telemedicine expansion can be particularly complex for pediatric patients, but approaches that satisfy privacy, security, and convenience will effectively increase pediatric enrollment capacity for telemedicine. Although telemedicine appears to be feasible and acceptable for clinical patients, questions about confidentiality, quality of care, and health disparities remain unanswered, so clinical guidelines are needed to guide best practices [21].

All races experienced a uniform increase in telehealth utilization, but the Black/African American community experienced the least in terms of percentage of increase. Research into racial disparities in telehealth by Rivera et al [22] in 2021 echoed these results by showing that African Americans are less likely to use telehealth and online services compared to Caucasians. In addition, according to Wegermann et al [23], there is still a gap in the growth of teletherapy relative to others for vulnerable populations, including those who are older, are non-Hispanic Black individuals, or have Medicare/Medicaid health insurance [23]. However, the results of this study are at odds with what Campos-Castillo and Anthony [24] found in that their data showed that African American respondents are more likely than Caucasians to report using telehealth because of the pandemic, particularly when perceiving the pandemic as a minor threat to their own health. Geographic differences could also have significant influences on racial disparities in telehealth utilization.

Although the Hispanic community experienced a large 42-fold increase, this was in part due to the fact that there was such low utilization (49 average monthly) during the prepandemic period. As for the percentage of the modality of visit type, telehealth was the lowest for Hispanics in all months examined except for April 2020, when they briefly passed the Black community by a small margin. As noted in multiple studies, telehealth access and utilization among Hispanics are inhibited by trust barriers, awareness, and eHealth literacy [25].

The willingness of younger patients to use telehealth, coupled with higher utilization rates among women, provides a positive glimpse into the future of telehealth. Racial and ethnic disparities in use due to many factors continue to exist, but the systemwide increase across all races/ethnicities also shows significant potential for sustained use. The types of services that the current set of available technologies can facilitate, however, provide insights into their usage rates.

When examining provider type, behavioral health and psychologists' services via telemedicine experienced the highest levels during the second phase, with over 60% of both services moving to the modality. Telehealth is well equipped to provide voice and video communications, which are some of the key aspects needed for these types of services. The next highest 2 were mental health services and therapists, a continued indication that the overall set of mental health services is most easily transferred to telemedicine. Dubin et al [26] found an almost 2-fold increase in the use of telemedicine by urologists, indicating that they have the ability to adopt and adapt telemedicine into their practice, but the barriers involved in the telemedicine technology itself still prevent many from taking advantage of it [26].

Although all services have diminished since their peak in the shutdown period of the second phase of the pandemic, behavioral health, psychologists' services, and mental health providers showed continued high utilization post-large-scale vaccine availability ranging from 10%-25%. All other services, including therapists, dropped back to below 5% telemedicine usage after vaccine availability. This is a possible indication that although telemedicine works for many services, in-person service is preferable to telemedicine. The study conducted by Smrke et al [27] reported the benefits of telemedicine for patients and clinicians in the long run since it can change cancer treatment delivery, particularly for patients with rare cancers who reside far away from expert centers.

Alhajri et al [28] found that video consultation should be frequently used in remote clinical consultation for acute conditions but that audio consultation is comparable in providing remote follow-up care for patients with chronic conditions and that audio consultation may greatly increase geographically accessible telemedicine services. Targeted efforts may be required during video visits to address patient populations that are older or have lower levels of knowledge [29]. At the same time, smartphone technology can serve as an extension of telemedicine, enabling the future of telehealth practices [30]. According to Orrange et al [31], patients' satisfaction using telemedicine is affected by their level of confidence in physicians and visit-related factors. The aims of improving access to treatment while avoiding overuse and fraud should be balanced in telemedicine policy, both in terms of regulation and in terms of payment [32].

In terms of rurality, a uniform increase in telehealth utilization was observed across highly rural, moderately rural, and urban counties, with moderately rural counties reaching the highest overall utilization rate. Research conducted by Chu et al [10] documented an increase in the adoption of telemedicine in rural and remote areas, but the use of telemedicine increased in urban and less rural populations during the COVID-19 pandemic. The observed tight correlation of telehealth trends across all types of counties indicates a significantly uniform willingness to use telehealth and overall consistency in the availability of telehealth. Studies such as Breton et al [33] indicate that mobility issues and patients living in remote areas could negatively impact telehealth utilization, but results from this analysis did not find significant access issues, even in the most rural counties. This could be attributed to the more widespread availability of high-speed mobile data services. To preserve the long-term viability of telemedicine programs in the aftermath of the COVID-19 pandemic, persuading third parties to continue to fund these services is necessary [34]. The future of telemedicine will also require addressing access barriers for vulnerable populations, such as people with disabilities, by making significant, long-term changes in technology, regulatory and legislative infrastructure, and customized solutions that meet the unique needs of patients and health systems [35].

Limitations

There are multiple factors that influence patient and practitioner willingness to use telemedicine, including connectivity, familiarity, and reimbursement policies, among others. The results of this study are likely applicable to states with similar demographics and rurality as Alabama, but broader applicability across the United States is questionable. However, Alabama has a high degree of rurality, poverty, and generally lower metrics in many socioeconomic factors, so if a technology such as telehealth can work in rural Alabama amongst its Medicaid population, then as long as connectivity exists elsewhere, we see no reasons, if policies allow, that similar results cannot be observed in other regions. In addition, it was reported that there were substantial differences in telemedicine completion rates among commercial insurance, Medicare, and Medicaid [36]. We only examined patients with a particular insurance type (ie, Medicaid) in this study, and researchers can also conduct similar analyses on telemedicine adoption of patients with other insurance types. Lastly, the study was descriptive in nature, which did not allow for the controlling of confounding factors. As a result, it is difficult to justify if any of the changes seen in a particular variable (ie, rurality) are not a product of other demographics (ie, if a particular demographic group of people may tend to live in a more rural area). Future studies may build multiple regression models to systematically investigate the factors that could influence telemedicine adoption.

Conclusion

The pandemic has had a jarring and brutal impact in terms of loss of life, economic stress and despair, and many other negative aspects. If a silver lining were to exist, though, from health care access Wand availability perspectives, it would be in the form of telemedicine adoption and utilization. Within the Alabama Medicaid population, telehealth services have proven to be extremely viable and have withstood the large-scale availability of vaccines to continue to be a modality of choice for health care in both rural and urban areas. Telemedicine is likely to continue to play an integral role in health care, and as a first step toward increasing the use of telemedicine, health care systems should focus on improving patient portal usage for better access to telemedicine services [37].

The potential benefits of sustained large-scale utilization of telehealth, especially in rural areas, are quite significant, yet concerns exist. As noted by Shachar et al [38], although telehealth may increase access, safety and privacy concerns are still common among users. In a similar vein, Bokolo [39] called for stakeholders and policymakers to confront the social, organizational, and technological determinants that are barriers to the increased adoption of telehealth. Although all of these concerns are valid and still exist, the pandemic triggered an interesting test of the telehealth systems nationwide, and the response was shown to be promising. Even in a state such as Alabama, with high rurality and high poverty, telehealth has not only shown to be an effective stop-gap measure but also has continued to show increased utilization postvaccine availability. The noted disparities among races with lower utilization rates among Black and Hispanic communities, coupled with the difference in usage amongst urban versus highly rural areas, stand as opportunities for increased focus in the future.

Acknowledgments

This work was supported in part by data and evaluation services provided by the Alabama Medicaid department. Although the findings were reviewed and confirmed by Alabama Medicaid, the results and views contained within are those only of the authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary Table S1.
[DOCX File , 90 KB - jmir_v24i7e38602_app1.docx]

Multimedia Appendix 2
Supplementary Table S2.
[DOCX File , 75 KB - jmir_v24i7e38602_app2.docx]

References

1. Ohannessian R, Duong TA, Odone A. Global telemedicine implementation and integration within health systems to fight the COVID-19 pandemic: a call to action. JMIR Public Health Surveill 2020 Apr 02;6(2):e18810 [FREE Full text] [doi: 10.2196/18810] [Medline: 32238336]
2. Monaghesh E, Hajizadeh A. The role of telehealth during COVID-19 outbreak: a systematic review based on current evidence. BMC Public Health 2020 Aug 01;20(1):1193 [FREE Full text] [doi: 10.1186/s12889-020-09301-4] [Medline: 32738884]
3. Kruse CS, Lee K, Watson JB, Lobo LG, Stoppelmoor AG, Oyibo SE. Measures of effectiveness, efficiency, and quality of telemedicine in the management of alcohol abuse, addiction, and rehabilitation: systematic review. J Med Internet Res 2020 Jan 31;22(1):e13252 [FREE Full text] [doi: 10.2196/13252] [Medline: 32012048]

4. Kichloo A, Albosta M, Dettloff K, Wani F, El-Amir Z, Singh J, et al. Telemedicine, the current COVID-19 pandemic and the future: a narrative review and perspectives moving forward in the USA. Fam Med Community Health 2020 Aug 18;8(3):e000530 [FREE Full text] [doi: 10.1136/fmch-2020-000530] [Medline: 32816942]
5. Alabama Medicaid. URL: https://medicaid.alabama.gov/ [accessed 2022-07-06]
6. Grigsby B, Brega AG, Bennett RE, Devore PA, Paulich MJ, Talkington SG, et al. The slow pace of interactive video telemedicine adoption: the perspective of telemedicine program administrators on physician participation. Telemed J E Health 2007 Dec;13(6):645-656. [doi: 10.1089/tmj.2007.0090] [Medline: 18092926]
7. Al Mohaya MA, Almaziad MM, Al-Hamad KA, Mustafa M. Telemedicine among oral medicine practitioners during COVID-19 pandemic and its future impact on the specialty. Risk Manag Healthc Policy 2021 Oct;Volume 14:4369-4378. [doi: 10.2147/rmhp.s325777]
8. Alabama GKI. Order of the State Health Officer Suspending Certain Public Gatherings Due to Risk of Infection by COVID-19. URL: https://www.alabamapublichealth.gov/legal/assets/soe-covid19-040320.pdf [accessed 2022-07-06]
9. Blue Cross and Blue Shield of Alabama Expands Telehealth Coverage as Customers? Health and Well-Being Remains Top Priority. URL: https://www.bcbsal.org/web/public/portal/pressreleaseinfo.html?articleId=97316005 [accessed 2022-07-06]
10. Chu C, Cram P, Pang A, Stamenova V, Tadrous M, Bhatia RS. Rural telemedicine use before and during the COVID-19 pandemic: repeated cross-sectional study. J Med Internet Res 2021 Apr 05;23(4):e26960 [FREE Full text] [doi: 10.2196/26960] [Medline: 33769942]
11. The COVID Tracking Project. URL: https://covidtracking.com [accessed 2022-07-06]
12. Alabama Coronavirus Vaccination Progress. URL: https://usafacts.org/visualizations/covid-vaccine-tracker-states/state/alabama [accessed 2022-07-06]
13. Analysis of Urban vs. Rural. URL: https://arhaonline.org/analysis-of-urban-vs-rural/ [accessed 2022-07-06]
14. Wu D, Gu H, Gu S, You H. Individual motivation and social influence: a study of telemedicine adoption in China based on social cognitive theory. Health Policy Technol 2021 Sep;10(3):100525. [doi: 10.1016/j.hlpt.2021.100525]
15. Muehlensiepen F, Knitza J, Marquardt W, Engler J, Hueber A, Welcker M. Acceptance of telerheumatology by rheumatologists and general practitioners in Germany: nationwide cross-sectional survey study. J Med Internet Res 2021 Mar 29;23(3):e23742 [FREE Full text] [doi: 10.2196/23742] [Medline: 33690147]
16. 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 2016 Oct 16;24(1):4-12. [doi: 10.1177/1357633x16674087]
17. Mann D, Chen J, Chunara R, Testa P, Nov O. COVID-19 transforms health care through telemedicine: evidence from the field. J Am Med Inform Assoc 2020 Jul 01;27(7):1132-1135 [FREE Full text] [doi: 10.1093/jamia/ocaa072] [Medline: 32324855]
18. Wood SM, White K, Peebles R, Pickel J, Alausa M, Mehringer J, et al. Outcomes of a rapid adolescent telehealth scale-up during the COVID-19 pandemic. J Adolesc Health 2020 Aug;67(2):172-178 [FREE Full text] [doi: 10.1016/j.jadohealth.2020.05.025] [Medline: 32611509]
19. Schweiberger K, Hoberman A, Iagnemma J, Schoemer P, Squire J, Taormina J, et al. Practice-level variation in telemedicine use in a pediatric primary care network during the COVID-19 pandemic: retrospective analysis and survey study. J Med Internet Res 2020 Dec 18;22(12):e24345 [FREE Full text] [doi: 10.2196/24345] [Medline: 33290244]
20. Patel P, Cobb J, Wright D, Turer RW, Jordan T, Humphrey A, et al. Rapid development of telehealth capabilities within pediatric patient portal infrastructure for COVID-19 care: barriers, solutions, results. J Am Med Inform Assoc 2020 Jul 01;27(7):1116-1120 [FREE Full text] [doi: 10.1093/jamia/ocaa065] [Medline: 32302395]
21. Barney A, Buckelew S, Mesheriakova V, Raymond-Flesch M. The COVID-19 pandemic and rapid implementation of adolescent and young adult telemedicine: challenges and opportunities for innovation. J Adolesc Health 2020 Aug;67(2):164-171 [FREE Full text] [doi: 10.1016/j.jadohealth.2020.05.006] [Medline: 32410810]
22. Rivera V, Aldridge MD, Ornstein K, Moody KA, Chun A. RESEARCHRacial and socioeconomic disparities in access to telehealth. J Am Geriatr Soc 2021 Jan 24;69(1):44-45 [FREE Full text] [doi: 10.1111/jgs.16904] [Medline: 33075143]
23. Wegermann K, Wilder JM, Parish A, Niedzwiecki D, Gellad ZF, Muir AJ, et al. Racial and socioeconomic disparities in utilization of telehealth in patients with liver disease during COVID-19. Dig Dis Sci 2022 Jan 28;67(1):93-99 [FREE Full text] [doi: 10.1007/s10620-021-06842-5] [Medline: 33507442]
24. Campos-Castillo C, Anthony D. Racial and ethnic differences in self-reported telehealth use during the COVID-19 pandemic: a secondary analysis of a US survey of internet users from late March. J Am Med Inform Assoc 2021 Jan 15;28(1):119-125 [FREE Full text] [doi: 10.1093/jamia/ocaa221] [Medline: 32894772]
25. King D, Khan S, Polo J, Solomon J, Pekmezaris R, Hajizadeh N. Optimizing telehealth experience design through usability testing in Hispanic American and African American patient populations: observational study. JMIR Rehabil Assist Technol 2020 Aug 04;7(2):e16004 [FREE Full text] [doi: 10.2196/16004] [Medline: 32749229]
26. Dubin JM, Wyant WA, Balaji NC, Ong WL, Kettache RH, Haffaf M, et al. Telemedicine usage among urologists during the COVID-19 pandemic: cross-sectional study. J Med Internet Res 2020 Nov 05;22(11):e21875 [FREE Full text] [doi: 10.2196/21875] [Medline: 33031047]

27. Smrke A, Younger E, Wilson R, Husson O, Farag S, Merry E, et al. Telemedicine during the COVID-19 pandemic: impact on care for rare cancers. JCO Global Oncol 2020 Nov(6):1046-1051. [doi: 10.1200/go.20.00220]
28. Alhajri N, Simsekler MCE, Alfalasi B, Alhashmi M, AlGhatrif M, Balalaa N, et al. Physicians' attitudes toward telemedicine consultations during the COVID-19 pandemic: cross-sectional study. JMIR Med Inform 2021 Jun 01;9(6):e29251 [FREE Full text] [doi: 10.2196/29251] [Medline: 34001497]
29. Fischer SH, Ray KN, Mehrotra A, Bloom EL, Uscher-Pines L. Prevalence and characteristics of telehealth utilization in the United States. JAMA Netw Open 2020 Oct 01;3(10):e2022302 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.22302] [Medline: 33104208]
30. Iyengar K, Upadhyaya GK, Vaishya R, Jain V. COVID-19 and applications of smartphone technology in the current pandemic. Diabetes Metab Syndr 2020 Sep;14(5):733-737 [FREE Full text] [doi: 10.1016/j.dsx.2020.05.033] [Medline: 32497963]
31. Orrange S, Patel A, Mack WJ, Cassetta J. Patient satisfaction and trust in telemedicine during the COVID-19 pandemic: retrospective observational study. JMIR Hum Factors 2021 Apr 22;8(2):e28589 [FREE Full text] [doi: 10.2196/28589] [Medline: 33822736]
32. Mehrotra A, Wang B, Snyder G. Telemedicine: What Should the Post-Pandemic Regulatory and Payment Landscape Look Like. URL: https://www.commonwealthfund.org/sites/default/files/2020-08/Mehrotra_Medicare_Telemedicine_ib.pdf [accessed 2022-07-06]
33. Breton M, Deville-Stoetzel N, Gaboury I, Smithman M, Kaczorowski J, Lussier M, et al. Telehealth in primary healthcare: a portrait of its rapid implementation during the COVID-19 pandemic. Healthc Policy 2021 Aug 10;17(1):73-90 [FREE Full text] [doi: 10.12927/hcpol.2021.26576] [Medline: 34543178]
34. 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]
35. 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]
36. Gmunder KN, Ruiz JW, Franceschi D, Suarez MM. Demographics associated with US healthcare disparities are exacerbated by the telemedicine surge during the COVID-19 pandemic. J Telemed Telecare 2021 Jun 23:1357633X2110259. [doi: 10.1177/1357633x211025939]
37. Gmunder KN, Ruiz JW, Franceschi D, Suarez MM. Factors to effective telemedicine visits during the COVID-19 pandemic: cohort study. JMIR Med Inform 2021 Aug 27;9(8):e27977 [FREE Full text] [doi: 10.2196/27977] [Medline: 34254936]
38. Shachar C, Engel J, Elwyn G. Implications for telehealth in a postpandemic future: regulatory and privacy issues. JAMA 2020 Jun 16;323(23):2375-2376. [doi: 10.1001/jama.2020.7943] [Medline: 32421170]
39. Anthony Jnr B. Implications of telehealth and digital care solutions during COVID-19 pandemic: a qualitative literature review. Inform Health Soc Care 2021 Mar 02;46(1):68-83. [doi: 10.1080/17538157.2020.1839467] [Medline: 33251894]

Abbreviations

IDA: Institute of Data and Analytics

Edited by C Basch; submitted 08.04.22; peer-reviewed by K Gmunder, F Shahzad; comments to author 28.04.22; revised version received 19.05.22; accepted 03.06.22; published 15.07.22.

Please cite as:
Xu P, Hudnall M, Zhao S, Raja U, Parton J, Lewis D
Pandemic-Triggered Adoption of Telehealth in Underserved Communities: Descriptive Study of Pre- and Postshutdown Trends
J Med Internet Res 2022;24(7):e38602
URL: https://www.jmir.org/2022/7/e38602
doi:10.2196/38602
PMID:35786564

Original Paper

Assessing Telemedicine Efficiency in Follow-up Care With Video Consultations for Patients in Orthopedic and Trauma Surgery in Germany: Randomized Controlled Trial

Jennifer Muschol[1], MSc; Martin Heinrich[2], Dr med; Christian Heiss[2], Prof Dr med Dr hc; Gero Knapp[2], Dr med; Holger Repp[1], Dr med; Henning Schneider[3], Prof Dr med; Ulrich Thormann[2], PD Dr med; Johanna Uhlar[3]; Kai Unzeitig[2], Dr med; Christian Gissel[1], Prof Dr

[1]Department of Health Economics, Justus Liebig University Giessen, Giessen, Germany

[2]Department of Trauma, Hand and Reconstructive Surgery, University Hospital Giessen, Giessen, Germany

[3]Institute of Medical Informatics, Justus Liebig University Giessen, Giessen, Germany

Corresponding Author:
Christian Gissel, Prof Dr
Department of Health Economics
Justus Liebig University Giessen
Klinikstraße 29
Giessen, 35392
Germany
Phone: 49 641 99 22070
Email: cg@phec.de

Abstract

Background: Telemedicine can help mitigate important health care challenges, such as demographic changes and the current COVID-19 pandemic, in high-income countries such as Germany. It gives physicians and patients the opportunity to interact via video consultations, regardless of their location, thus offering cost and time savings for both sides.

Objective: We aimed to investigate whether telemedicine can be implemented efficiently in the follow-up care for patients in orthopedic and trauma surgery, with respect to patient satisfaction, physician satisfaction, and quality of care.

Methods: We conducted a prospective randomized controlled trial in a German university hospital and enrolled 60 patients with different knee and shoulder conditions. For follow-up appointments, patients received either an in-person consultation in the clinic (control group) or a video consultation with their physician (telemedicine group). Patients' and physicians' subsequent evaluations of these follow-up appointments were collected and assessed using separate questionnaires.

Results: On the basis of data from 52 consultations after 8 withdrawals, it was found that patients were slightly more satisfied with video consultations (mean 1.58, SD 0.643) than with in-clinic consultations (mean 1.64, SD 0.569), although the difference was not statistically significant (P=.69). After excluding video consultations marred by technical problems, no significant difference was found in physician satisfaction between the groups (mean 1.47, SD 0.516 vs mean 1.32, SD 0.557; P=.31). Further analysis indicated that telemedicine can be applied to broader groups of patients and that patients who have prior experience with telemedicine are more willing to use telemedicine for follow-up care.

Conclusions: Telemedicine can be an alternative and efficient form of follow-up care for patients in orthopedic and trauma surgery in Germany, and it has no significant disadvantages compared with in-person consultations in the clinic.

Trial Registration: German Clinical Trials Register DRKS00023445; https://www.drks.de/drks_web/navigate.do?navigationId=trial.HTML&TRIAL_ID=DRKS00023445

(J Med Internet Res 2022;24(7):e36996) doi:10.2196/36996

KEYWORDS

telemedicine; video consultations; follow-up; efficiency; orthopedic; trauma surgery; mobile phone

Introduction

Background

International health care systems are facing several major challenges. Some of these challenges are structural and have evolved over the years, while others have occurred as sudden shocks. Demographic change and a shortage of health care professionals are among the increasingly important structural challenges and have been impacting patient care for years. On the one hand, the rising number of older and multimorbid patients is leading to a demographic change, which is associated with a higher demand for health care services. On the other hand, there is a growing shortage of specialists to meet this demand efficiently. Simultaneously, an asymmetrical distribution of medical service providers leads to deficits in health care. Particularly in rural regions, patients have to travel longer distances, and thus incur higher costs. In the long term, this could lead to limited access to health care for a subset of patients [1-3].

Beyond these structural challenges, health care systems have recently had to cope with the sudden COVID-19 pandemic [4], starting with its global outbreak in March 2020 [5]. The pandemic has placed several important restrictions on the delivery of medical care; for example, social distancing has become necessary to avoid infections [6,7]. Hospitals, which are at a particular risk of causing a pandemic outbreak owing to their high number of interactions and patients, have introduced protective measures [8,9]. Some patients avoid medical appointments for fear of infection [10,11], and hospitals have been postponing nonurgent treatments and interventions to save resources [6,7,12]. The lack of physician-patient interactions and avoidance of treatments could lead to worsening health outcomes in the future [13].

Both structural challenges and the pandemic shock will likely have a long-term impact on the health care system and delivery of care. As a result, existing structures will need to be reconsidered [2,3,14].

One important tool for overcoming these challenges and guaranteeing effective health care in the medium to long term could be the use of telemedicine. Telemedicine offers the ability to provide medical care through real-time video consultations, without the need for personal contact and regardless of location. This could free up clinical resources, improve access to care, and increase safety for patients and medical staff [3,15-18].

Telemedicine is already being applied successfully in various medical fields [19,20], but its use has so far been less common in orthopedic and trauma surgery owing to the specialty's heavy reliance on palpation and dynamic testing and telemedicine's inherent constraints [7]. In addition, regardless of medical specialty, there were barriers that still negatively impacted readiness for adoption despite the benefits of telemedicine. These barriers included, for example, resistance to change, lack of technical literacy, or uncertainties about costs and reimbursement [21,22]. However, since the outbreak of the COVID-19 pandemic, the need for telemedicine has increased considerably in the field of orthopedic and trauma surgery, among other medical areas [7,16,18]. More specifically, telemedicine can support outpatient care in hospitals, such as follow-up examinations to prior interventions [23]. While vital for successful treatment [24], these follow-up examinations entail a travel burden for patients who are often immobile or in pain due to their condition. Therefore, telemedicine could offer a suitable alternative [25].

In 2019, German hospitals admitted a total of 854,410 patients in orthopedic surgery and 759,356 patients in trauma surgery, making the combination of orthopedic and trauma surgery one of the largest areas of care in Germany [26]. Increasing the use of telemedicine to relieve clinics and patients of unnecessary burdens in this broad field could provide significant benefits. Although these benefits can be determined only by clinical evaluation, randomized controlled trials (RCTs) examining telemedicine in orthopedic and trauma surgery are rare, with few exceptions.

In an RCT, Buvik et al [23,27,28] compared standard consultations in an orthopedic outpatient clinic of a hospital with video consultations assisted by a trained nurse at a regional medical center in Norway. It was shown that telemedicine is a safe alternative, that its use can be cost-effective, and that there are no significant differences in patient satisfaction and health status between the treatment group and the control group [23,27,28].

Sathiyakumar et al [29] also found no significant differences in patient satisfaction between telemedicine and in-person follow-up for patients with closed orthopedic trauma injuries in a level 1 trauma center in the United States. In this RCT, telemedicine was associated with time and travel savings for patients [29].

The use of telemedicine for a postoperative follow-up of arthroscopic rotator cuff repair surgery was investigated by Kane et al [30] in the United States. Their RCT concluded that telemedicine can be used safely and effectively for this condition, that patient satisfaction was similar, and that time savings were achieved for both patients and physicians [30].

Objectives

However, prior research has left several questions unaddressed, which are considered based on our research design. One of them is whether the use of telemedicine is efficient not only for a restricted number of individual diseases but also for a wider range of medical conditions. Another important question is with regard to the practicality of telemedicine without the need to involve additional staff to assist patients during video consultations [23,27,28]. Furthermore, it is questionable whether international study results can be transferred to the German health care system, especially because studies show that German patients are skeptical about the use of telemedicine [31].

The aim of our RCT was to investigate whether telemedicine can be used efficiently in follow-up care for patients in orthopedic and trauma surgery in Germany. To answer this question, the RCT compared an in-person consultation in a German university hospital (level 1 trauma center) with the use of telemedicine, namely a video consultation between the physician and patient. All consultations were for the follow-up

care of patients with knee and shoulder conditions who displayed a variety of conditions, had previously been treated in the clinic, and were eligible to participate in the study. For their video consultation, patients did not have to travel to the clinic but could have their follow-up appointment on the web, regardless of their location. For this study, the aspects of patient satisfaction, physician satisfaction, and quality of care were considered as the most important factors to quantify the output of telemedicine. Therefore, they were included in the evaluation of telemedicine under the overarching term "efficiency." It is hoped that studying telemedicine in broad-based use for follow-up care and analyzing its effects comprehensively will contribute to informing health care providers' decision-making in future.

Methods

Study Design

This study was conducted as an open, prospective, interventional, 1:1 randomized controlled monocenter trial at a German university hospital (University Hospital Giessen, Department of Trauma, Hand and Reconstructive Surgery). The randomized and controlled design is based on the CONSORT (Consolidated Standards of Reporting Trials) [32]. The effects of telemedicine on follow-up care were examined with the parallel implementation of an intervention group, which received follow-up care through a real-time video consultation, and a control group, which received a standard follow-up consultation in the clinic.

Ethics Approval

The local ethics committee of the University of Giessen reviewed and approved this study (AZ 73/20). The study was registered in the German Clinical Trials Register (ID: DRKS00023445).

Definition and Characteristics of the Trial Population

The trial population consisted of knee and shoulder patients who have already been treated in the department. The patients' medical conditions varied, and Multimedia Appendix 1 lists their ICD-10 codes. Their medical conditions included, for example, fractures of the patella and femur, impingement syndrome of the shoulder, and orthopedic joint implants.

To adequately guarantee the safety of patient care, recruitment for the RCT observed the following inclusion criteria in addition to the ICD-10 codes: (1) patients need the ability to consent, as well as the mental and physical ability to participate in the telemedical consultation. (2) As part of the consultation, patients' conditions should require no more than a visual examination and a conversation without the need to be touched or moved by the treating physician or other physical interactions. For legal reasons, (3) a previous outpatient or inpatient stay at the clinic is required, and (4) patients must be ≥18 years. To be able to use telemedicine, it is assumed that (5) patients own a computer, laptop, tablet, or smartphone, including a microphone and camera, and (6) they have a stable internet connection. Finally, (7) patients have to speak German to understand the declaration of consent.

The concern with patient safety is also reflected in the exclusion criteria. Thus, patients with (1) neurological diseases that do not allow the use of computer systems and (2) patients with a diagnosis of dementia, blindness, or deafness were excluded. In addition, patients were excluded if they (3) have a need for in-person presence and on-site diagnostics or treatments (eg, medical imaging, laboratory, stitches, or drainage) or (4) have to be touched or moved by the treating physician. This ensured that patients who required personal contact with a physician were not at risk. Finally, (5) a lack of willingness to participate in the study or (6) the failure to consent were further added to the exclusion criteria.

Recruitment and Randomization of Study Participants

After the initial screening for inclusion and exclusion criteria, patients were asked either at the clinic or by telephone if they would like to participate in the study during their next follow-up appointment. To be able to participate, the patients had to provide informed consent after receiving written and oral information. Consent could be withdrawn at any time, without providing reasons.

We followed a 2-armed parallel group design, and patients enrolled in the study were randomly assigned in a 1:1 ratio to either the intervention arm (telemedicine follow-up) or the control arm (in-person follow-up consultation in the clinic). To ensure better balance between the arms while minimizing predictability, block randomization with randomly selected block sizes of 4, 6, and 8 was applied [32]. One member of the study staff organized the allocation of the blocks, while a different member performed the randomization of the patients. For this purpose, sealed envelopes were used. Given that the intervention was a video consultation, blinding of the physicians or patients was not possible. At the end of the recruitment process, depending on the treatment arm, the patients received an appointment either in the clinic or for a video consultation. Patients in the intervention group also received written instructions on how to be prepared for the video consultation to minimize potential technical difficulties.

Procedures

Intervention Arm

Study participants who were assigned to the intervention arm received a one-time telemedical follow-up via a real-time videoconference instead of a standard consultation in the department. The one-time appointment was intended to avoid bias through learning effects. The web-based video consultation used the web-based software CLICKDOC of the German telemedicine provider CGM Mobile Services GmbH. This software is certified for and widely used in the German health care system. On the day of their appointment, the patients received log-in details for the video consultation from their physicians via SMS text message or email. No additional registration was required for the video consultation. At the arranged time, both the attending physician and the patient logged in to the software program, and the video consultation was conducted. Patients were able to use a computer, laptop, tablet, or smartphone to join the video consultation. If technical problems were noted, patients were contacted by phone and

scheduled for a clinic visit, as needed. Immediately after the video consultation, the patients received log-in details via email and were asked to evaluate the consultation via web-based questionnaires.

Control Arm

Study participants, who were assigned to the control group, attended a standard follow-up consultation at the university hospital. This follow-up was conducted by the same physicians who also treated the intervention arm. Immediately after the consultation, patients in the control arm were asked to fill out the questionnaires at the clinic.

Outcome and Data Collection

To answer the research questions and analyze the outcome parameters, that is, patient satisfaction, physician satisfaction, and quality of care, different questionnaires were completed by patients and physicians.

The primary outcome patient satisfaction was measured using the German questionnaire Zufriedenheit in der ambulanten Versorgung–Qualität aus Patientenperspektive of the National Association of Statutory Health Insurance Physicians in Germany [33,34]. This validated questionnaire is an appropriate way of investigating patient satisfaction in the German outpatient sector. To adequately reflect the specific conditions of this study, individual items of the questionnaire were modified and some were added. This included excluding questions that were not relevant to the purpose of our study. The additional questions addressed whether patients experienced the treatment they wanted, how much time the physician had provided, how comfortable patients felt with the treatment, whether they were agitated, and how punctual the treatment appointment was. In addition, patients were asked to rate their overall satisfaction with their respective follow-up appointment using German school grades, where grade 1 represents "very good" and grade 6 represents "inadequate." The other questions were answered using a 4-point Likert scale, where higher scores represented higher satisfaction. Finally, the patients were asked which option they would choose for their next follow-up appointment.

Physician satisfaction, as one of the secondary outcome parameters, was assessed by questionnaires that the physicians answered following each patient consultation. The questionnaires were self-designed and differed slightly depending on the study arm. In both groups, the physicians were asked whether all medical questions could be clarified, which option the physicians would choose for the next follow-up appointment, and how satisfied they were with the consultation. Satisfaction was also surveyed using school grades in this case. The questionnaire in the intervention group was supplemented with the questions of whether a technical irregularity occurred during the treatment and whether the consultation had to be terminated due to this malfunction.

To evaluate the quality of care as a secondary outcome, patients received the German version of the "EQ-5D-5L" questionnaire from the EuroQol Group during enrollment [35]. Patients were asked to rate their current health-related quality of life between 0 and 100 on a visual analog scale (VAS). After 3 months, the questionnaire was completed again to measure the impact of the interventions on health-related quality of life.

Sample Size

We performed a priori power analysis using the software G*Power 3.1.9.6 (Heinrich Heine University) which calculates the sample size based on the power, significance level, and effect size [36]. To determine the effect size, we used the study by Sharareh and Schwarzkopf [25] as a baseline, which conducted a group comparison of patient satisfaction with telemedicine. As the resulting effect size represented a very strong effect that we did not expect in our study, we used half of the effect size (1.095) to perform the sample size calculation. This resulted in approximately 19 study participants for each group to achieve a power of 90% in a 2-sided *t* test for independent samples with a global significance level of 5%. The sample size was increased by 10% for both groups to accommodate potential dropouts or withdrawals and by another 10% to counteract a potentially skewed distribution of patient satisfaction. This resulted in a case number of 23 patients per randomization arm. To consider the possible loss of power when using nonparametric methods, the sample size was increased to 30 patients per arm, resulting in a total of 60 enrollments.

Statistical Analysis

The statistical evaluation of the study included descriptive statistics of the demographic characteristics and parameters collected in the questionnaires. Continuous and ordinal data were reported as mean values, SDs, and medians. Categorical data were presented as absolute and relative frequencies. Differences between the 2 study arms were analyzed using the Mann-Whitney *U* test or Fisher exact test, and effect sizes were reported by Pearson correlation coefficient (*r*) or Cramer V. These nonparametric tests were used because most of the data were not normally distributed, assumptions were not met, or the underlying scale was ordinal. For patient and physician satisfaction, a subgroup analysis based on medical indications was performed. In addition, the Wilcoxon signed-rank test was used to evaluate the longitudinal data of the EQ-5D-5L VAS. Owing to incomplete questionnaires, the reported group size (n) was different for each test. The data were analyzed based on intention-to-treat. The *P* value was set a priori at .05 to test 2-sided significance. The Bonferroni-Holm correction was applied but did not affect the reported results.

To examine the suitability of telemedicine for follow-up appointments in detail and to investigate the type of patients who would use telemedicine, a binary logistic regression was performed. The variable "Which option would you choose for your next appointment?" from one of the patient questionnaires was used as the dependent variable, with the dichotomous outcome "telemedical follow-up" or "standard consultation." The independent variables added to the model were the categorical parameters "group" with the outcome telemedicine group or control group; "indication" in the form of knee or shoulder; "sex" as male or female; "age" divided into the categories 18 to 40 years, 41 to 60 years, and >60 years; and finally, prior experience with video calls. This exploratory model sought to investigate the factors that influence the decision to use telemedicine when offering video consultations in clinical

practice. Bootstrap validation was performed to confirm the validity of the results. A receiver operating characteristic curve was used to assess the accuracy of the model.

Results

Overview

The patients were recruited and attended their follow-up appointments between September 2020 and April 2021. The last questionnaires for the second data collection of the EQ-5D-5L were sent in July 2021. For organizational reasons, the number of eligible patients could not be recorded until 2 months after the start of recruitment, resulting in a total of 102 eligible patients.

In total, 60 patients agreed to participate in the study and were randomized; 30 patients were allocated to the intervention arm and 30 patients, to the control arm. After randomization, 8 patients withdrew from the study. None of these patients were excluded by the physicians. Thus, 26 patients in the intervention arm and 26 patients in the control arm could be analyzed. Figure 1 shows the CONSORT flow diagram outlining the process of patient recruitment and data analysis. In total, 100% (26/26) of patients in the intervention arm and 90% (26/29) of patients in the control arm completed the questionnaires after the follow-up appointment. With regard to the physician questionnaires, 100% (26/26) in the telemedicine group and 96% (25/26) in the control group were completed. In the intervention group, 100% (26/26) of the EQ-5D-5L questionnaires were returned at baseline, and 69% (18/26) were returned after 3 months; in the control group, 88% (23/26) of the questionnaires were returned at baseline, and 58% (15/26) were returned after 3 months.

Demographic characteristics of patients, such as sex, age, medical indication, distance from clinic, and health status showed no significant differences between the 2 groups (Table 1).

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram.

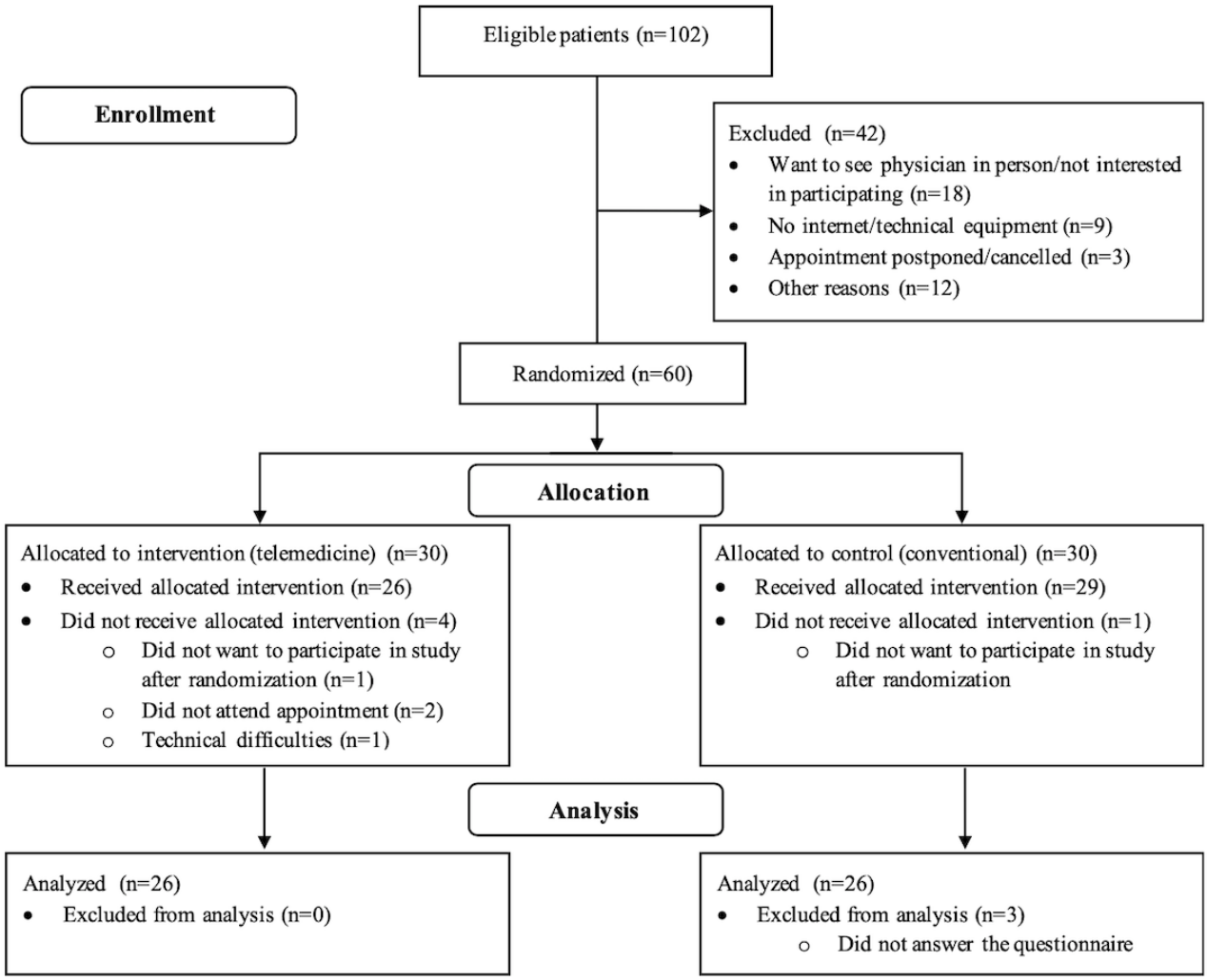


Table 1. Demographic characteristics of patients.

	Telemedicine group (n=26)	Control group (n=26)	*P* value
Sex, n (%)			.99[a]
Female	11 (42)	10 (39)	
Male	15 (58)	16 (61)	
Age (years), n (%)			.36[a]
18-40	7 (27)	5 (19)	
41-60	17 (65)	15 (58)	
>61	2 (8)	6 (23)	
Medical indication, n (%)			.99[a]
Knee	10 (39)	9 (35)	
Shoulder	16 (61)	17 (65)	
Distance from clinic (km), mean (SD)	37.00 (32.06)	31.58 (22.62)	.65[b]
Self-assessed health status, mean (SD)	2.88 (1.033)	2.91 (0.848)	.96[b]

[a]Fisher exact test.

[b]Mann-Whitney *U* test.

Patient Satisfaction

To measure perceived patient satisfaction, patients in both groups were asked to rate their satisfaction with their respective follow-up appointment using German school grades.

Although group comparison showed that patients were slightly more satisfied with telemedicine follow-up (mean 1.58, SD 0.643) than with in-person follow-up in the clinic (mean 1.64, SD 0.569), the difference was not statistically significant (P=.69; Table 2). This result was not affected by a subgroup analysis of the 2 medical indications, namely knee or shoulder. Analysis of the other aspects of the adapted Zufriedenheit in der ambulanten Versorgung–Qualität aus Patientenperspektive questionnaire, such as organization, information, interaction, and participation, showed no significant differences between the groups, with a few exceptions. The waiting time (P<.001), atmosphere (P<.001), and punctuality of the appointment (P=.002) were more satisfying for patients in the telemedicine group than in the control group, with medium to strong effects (r=0.440 to r=0.760). Box and whisker plots of all distributions can be found in Multimedia Appendix 2.

A strong effect was also evident in the preference for the next follow-up appointment between the groups, which was analyzed with Fisher exact test (V=0.542). While patients in the control group preferred to visit the clinic again (16/25, 64%), almost all patients in the telemedicine group (23/26, 88%) chose telemedicine for their next follow-up appointment (P<.001). However, a clear majority (32/51, 63%) of all patients chose a video consultation for their next appointment, whereas only 37% (19/51) chose a standard consultation.

Table 2. Patient satisfaction.

	Telemedicine group			Control group			*P* value[a]	Pearson correlation coefficient (*r*)
	Value, n	Value, mean (SD)	Value, median	Value, n	Value, mean (SD)	Value, median		
Overall patient satisfaction	26	1.58[b] (0.643)	1.5	25	1.64[b] (0.569)	2.00	.69	0.071
Satisfaction knee patients	10	1.80[b] (0.632)	2.00	9	1.89[b] (0.601)	2.00	.95	0.077
Satisfaction shoulder patients	16	1.44[b] (0.629)	1.00	16	1.50[b] (0.516)	1.5	.72	0.092
How satisfied are you with the waiting time?	26	2.88[c] (0.326)	3.00	25	2.12[c] (0.881)	2.00	<.001	0.546
How satisfied are you with the atmosphere?	26	2.85[c] (0.368)	3.00	25	2.08[c] (0.277)	2.00	<.001	0.760
How punctual was your appointment?	26	2.35[c] (0.689)	2.00	25	1.60[c] (0.913)	2.00	.002	0.440

[a]Mann-Whitney *U* test.

[b]German school grades; 1=very good to 6=inadequate.

[c]4-point Likert scale; higher scores=higher satisfaction.

Physician Satisfaction

Physicians in the control group were significantly more satisfied with the follow-up appointments (mean 1.32, SD 0.557) than those in the telemedicine group (mean 2.42, SD 1.419; *P*=.001; *r*=0.466), as shown in Table 3. The subgroup analysis showed that this difference was also significant for the treatment of shoulder patients (*P*=.006) but not for knee patients (*P*=.08). However, a further group comparison, in which video consultations with technical irregularities were removed, revealed no significant group differences in physician satisfaction (mean 1.47, SD 0.516 and mean 1.32, SD 0.557; *P*=.31). In addition, there were no significant differences in their ability to address all relevant medical questions (telemedicine group: 25/26, 96%; control group: 25/25, 100%; *P*=.99; *V*=0.139).

Table 3. Physician satisfaction.

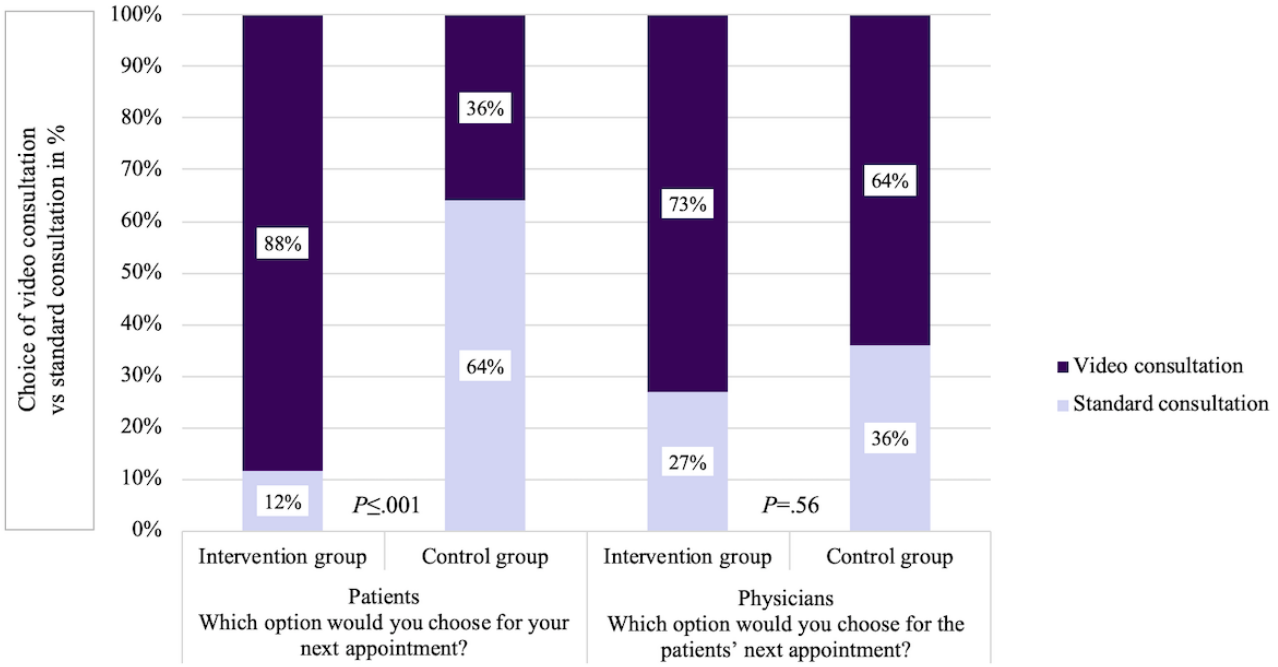
	Telemedicine group			Control group			*P* value[a]	Pearson correlation coefficient (*r*)
	Value, n	Value, mean (SD)	Value, median	Value, n	Value, mean (SD)	Value, median		
Overall physician satisfaction	26	2.42[b] (1.419)	2.00	25	1.32[b] (0.557)	1.00	.001	0.466
Satisfaction knee patients	10	2.30[b] (1.829)	1.50	10	1.10[b] (0.316)	1.00	.08	0.449
Satisfaction shoulder patients	16	2.50[b] (1.155)	2.00	15	1.47[b] (0.640)	1.00	.006	0.492
Physician satisfaction without technical irregularities	15	1.47[b] (0.516)	1.00	25	1.32[b] (0.557)	1.00	.31	0.167

[a]Mann-Whitney *U* test.

[b]German school grades; 1=very good to 6=inadequate.

For the next follow-up appointment, physicians recommended a telemedical consultation for most patients, regardless of the study arm (19/26, 73% of the telemedicine group; 16/25, 64% of the control group; *P*=.56; *V*=0.098). Overall, physicians recommended telemedicine to 69% (35/51) of patients for further follow-up. A comparison of the patients and physicians and their respective choices for the next follow-up appointment is shown in Figure 2.

Figure 2. Patient and physician choice of next follow-up appointment.



Quality of Care

Quality of care was assessed by surveying patients' perceived health-related quality of life before and after the follow-up visit, using the EQ-5D-5L VAS. As shown in Table 4, the differences in quality of life between the groups were not significant, neither at baseline (*P*=.24) nor after treatment (*P*=.69). The difference in quality of life before and after the follow-up appointment was also not statistically significant between the intervention and control arms (*P*=.19). In this case, the group size changed because only complete data sets could be considered for analysis. In both groups, it was shown with the Wilcoxon signed-rank test that the perceived average quality of life increased after treatment, although not significantly, (telemedicine group: mean 69.77, SD 20.551 to mean 70.44, SD 19.509; *P*=.93; *r*=0.027; control group: mean 66.30, SD 18.292 to mean 69.33, SD 15.216; *P*=.11; *r*=0.440).

Table 4. Quality of care.

	Telemedicine group			Control group			*P* value[a]	Pearson correlation coefficient (*r*)
	Value, n	Value, mean (SD)	Value, median	Value, n	Value, mean (SD)	Value, median		
EQ-5D-5L VAS[b] baseline	26	69.77[c] (20.551)	79.00	23	66.30[c] (18.292)	75.00	.24	0.169
EQ-5D-5L VAS 3 months	18	70.44[c] (19.509)	72.5	15	69.33[c] (15.216)	75.00	.69	0.073
Δ EQ-5D-5L VAS	18	0.833[c] (16.468)	0.00	14	6.43[c] (13.921)	6.00	.19	0.237

[a]Mann-Whitney *U* test.

[b]VAS: visual analog scale.

[c]Scale from 0 to 100.

Binary Logistic Regression

Binary logistic regression was used to examine the types of patients who would use telemedicine for their follow-up appointment. For this purpose, the potential influence of different variables on patients' preference for their next follow-up appointment was analyzed (Table 5). The model was statistically significant (χ^2_6=22.3; *P*=.001; Nagelkerke R^2=48.4%). It was found that medical indication, sex, nor age had a significant influence on the choice of telemedicine. However, previous experience with video calls before the study (*P*=.03) and the respective study arm in which the patients were treated (*P*=.001) contributed significantly to predicting the choice of telemedicine.

Table 5. Binary logistic regression.

Variables	Coefficient (β; SE)	*P* value	Odds ratio (95% CI)
Study arm	2.760 (0.836)	.001	15.793 (3.066-81.351)
Indication	.004 (0.810)	.99	1.004 (0.205-4.913)
Sex	−0.686 (0.868)	.43	0.504 (0.092-2.761)
Previous experience with video calls	1.726 (0.802)	.03	5.620 (1.168-27.038)
Age (years)			
18-40	−0.388 (1.280)	.76	0.678 (0.055-8.333)
41-60	.587 (1.044)	.57	1.799 (0.232-13.919)
>61 (reference)	__[a]	.59	—
Constant	−1.436 (1.031)	.16	0.238[b]

[a]Reference category.

[b]95% CI value is not applicable.

Patients were 15.8 times more likely to consider telemedicine as a treatment option for further follow-up care if they had already experienced telemedicine than if they had previously been in the control group. Prior experience with general videoconferencing increased the likelihood of participating in telemedicine by 5.6-fold compared with no experience. As a measure of accuracy, the area below the receiver operating characteristic curve was 0.86 (P<.001), which indicated that the model has an appropriate fit to predict whether a patient will choose telemedicine for the next follow-up appointment.

Discussion

Principal Findings

This study aimed to investigate whether telemedicine can be used efficiently for outpatient orthopedic and trauma surgery follow-up care in Germany in a university hospital from the perspective of patients, physicians, and the quality of care. Our data analysis showed that the use of telemedicine had no significant drawbacks compared with traditional clinical consultations in almost all aspects studied. Patients were even slightly more satisfied with telemedicine, regardless of their medical condition, although the difference was not statistically significant. In addition to overall satisfaction, the authors analyzed specific indicators of satisfaction. Aspects such as waiting time, atmosphere, and punctuality can be significantly improved by using telemedicine. Advantages for patients might arise from the fact that they can interact with their physicians in a familiar environment and do not have to be present in the hospital. The atmosphere, which is the sentiment that patients experience during medical consultations, is perceived to be more pleasant at home than at the hospital. In addition, the waiting time and punctuality of consultations could be evaluated more positively, as patients could use the time at home and are not limited to waiting in the clinic waiting room. In addition, being able to consult a physician from the comfort of home without having to travel and without experiencing long waiting times is a benefit that could have a positive impact on patients' well-being. The results in the RCTs by Buvik et al [28], Sathiyakumar et al [29], and Kane et al [30] are similar to our findings regarding patient satisfaction. However, these RCTs differ from our study design, as they focus mainly on using telemedicine for individual medical conditions or in a specific setting, such as an outpatient clinic with staff support [28-30].

Although the group comparison showed that physicians were less satisfied with telemedicine than with standard consultations, this lower satisfaction can probably be attributed to technical irregularities. When they were excluded from the analysis, there was no significant difference in satisfaction between the 2 groups for physicians as well. The comparison also showed that technical difficulties have a stronger influence on physician satisfaction than on patient satisfaction. This could be related to the fact that physicians must follow a fixed schedule, which is sensitive to disruptions. The fact that patients are more satisfied with telemedicine than physicians was also identified in the study by Buvik et al [28]. Nevertheless, the high number of video consultations with technical irregularities (11/26, 42%) could have a negative impact on satisfaction with telemedicine over time, and we could not determine whether the irregularities were system related or because of human error. This challenge could be mitigated by the fact that ongoing technological improvements could help make telemedical consultations both easier and more reliable.

Patients' health-related quality of life did not differ significantly between groups. This might indicate that the application of telemedicine is suitable for the patient group studied and does not have a negative impact on the quality of care. However, it must be noted that telemedicine is suitable only for patients who are already in an advanced stage of the treatment process and who do not currently require follow-up care in the clinic. Therefore, the disease pattern and condition of each patient were reviewed by the physicians before their participation in the study. Generally, the change in quality of life will be less pronounced at this later stage of the treatment process.

All patients in the telemedicine group of this study experienced only 1 telemedical consultation. However, even with this minimal gain in experience, it can be seen that these patients would more frequently opt for telemedicine than those in the control group. Results of other RCTs showed the same conclusion [28-30,37]. Kane et al [30] argued that this could

be associated with the fact that people prefer the known rather than the unknown. Thus, familiarity could influence the choice of the type of consultation. However, in addition to this effect, we assume that there is an initial barrier for patients to use telemedicine, for example, technical hurdles. This barrier is overcome for the vast majority of patients once they have participated in their first video consultation. Thereafter, patients were more willing to use telemedicine again, indicating that they considered it an appropriate treatment option. Therefore, it is important to introduce eligible patients to the use of telemedicine and to support them in case of potential uncertainty. At the same time, the learning effect could also increase long-term satisfaction with telemedicine as patients become more confident in using the digital application and learn how to avoid sources of error. Therefore, satisfaction may be even higher among patients who use telemedicine more often. In contrast, physicians recommend telemedicine for most patients, regardless of the study arm. This may be because they prioritize medical value over patients' prior experiences with this consultation format. The fact that physicians recommend telemedicine despite their initially lower level of satisfaction with it is a further indication of the appropriateness of telemedicine.

Unlike other studies, we also investigated for which patients telemedicine is best suited, as the sensible choice of a promising target group is crucial for the success of telemedicine applications in practice. We were able to show that telemedicine need not be restricted to a specific group of patients but can be provided broadly. Telemedicine was positively evaluated by both knee and shoulder patients with varying ICD-10 codes. Furthermore, binary logistic regression revealed that demographic characteristics had no significant influence on the choice of telemedicine; only prior experience was decisive. This is particularly relevant for clinical practice, as in the long run only the broad-based use of telemedicine for a heterogeneous group of patients is likely to be efficient. When treating single conditions or a small subset of patients, important economies of scale might not be achieved. However, it should be noted that the patients in our trial were comparatively young. Thus, we cannot reject with certainty that a particularly older age might reduce patients' willingness to participate in telemedicine. However, with the rapid progress of digitalization and its use, this would become less relevant in the future [38].

Nevertheless, the expanded adoption of telemedicine is accompanied by barriers for certain patient groups. In some cases, the use of digital technologies is restricted to older adults, socially deprived people who lack financial means, people without internet access (eg, in rural areas), or members of ethnic minorities [39]. To prevent potential disadvantages and exclusion of these patient groups, policy makers need to consider the following aspects: national internet access infrastructure, access to digital equipment, availability of digital applications in the required languages, deployment of health workers to support patients during video consultations, access to trainings on how to use telemedicine, and the introduction of programs that support digital health literacy [40].

Limitations

Our study had some limitations. First, we based our sample size calculation on a study by Sharareh and Schwarzkopf [25], which measured a large difference in satisfaction between groups. As a result, our recruited sample size consisted of only 60 patients, which corresponds to a larger expansion of the original sample size calculated. On the basis of our data, we could not detect such a large difference between the groups. Thus, the restricted sample size may have influenced the statistical power of the tests. For example, the results of the binary logistic regression could become more robust with a larger sample size. Furthermore, because of the small sample size, we had to validate the binary logistic regression by bootstrapping and could not split the data set to perform a separate evaluation and validation. Nevertheless, our sample size was comparable with that of other studies, such as that of Kane et al [30].

Another limitation concerns the questionnaires used. International studies have shown that the number of validated questionnaires in this context is limited [23]. This problem is particularly acute in German studies. Therefore, we had to adapt validated questionnaires and partly create them.

The use of pen-and-paper questionnaires, on the one hand, and web-based questionnaires, on the other hand, could also have led to discrepancies. In particular, all questions had to be completed in the web-based questionnaires, but this was not true for the pen-and-paper questionnaires completed in the clinic. However, for organizational reasons, no uniform implementation was possible. This problem also arose for comparable studies. On the other hand, studies have shown that patients usually provide similar health-related answers regardless of survey formats [41-43].

When evaluating the results, it should be noted that all patients recruited from the intervention and control groups consented to participate in telemedicine. Thus, there was an initial interest in telemedicine among participants. This might have led to a self-selection bias in favor of higher satisfaction with telemedicine from the start because patients who were more comfortable with digitalization were more likely to participate in the study [22]. Although all patients consented to undergo a video consultation, it was found that patients in the intervention group were more likely to choose a video consultation for their next follow-up appointment than those in the control group, further supporting the effect of comfortability. Although this self-selection bias is evident for all telemedicine evaluations with a similar study design, it leads to the limitation that the data do not show results for the general population but only show results for patients with a baseline interest in telemedicine [22]. Short of forcing patients to participate in telemedicine, a procedure that appears both unethical in principle and unfeasible in clinical practice, there is no acceptable way of addressing this limitation. In 2018, the percentage of German patients who found video consultations helpful in orthopedic and trauma surgery was 30.5% [31]. However, the higher willingness to participate in our study might suggest that this number will increase in the long term, making our results more generalizable.

Finally, we considered only 1 follow-up consultation in our study design to avoid bias owing to potential learning effects.

Thus, no conclusions regarding long-term satisfaction with telemedicine can be made in the context of our study. We suggest that future studies analyze long-term satisfaction with telemedicine in orthopedic and trauma surgery follow-up care. In this context, it should also be investigated how challenges in standard clinic appointments, such as undetected diseases or complications, develop in the context of performing video consultations, particularly in larger patient populations. Future studies concerning the acceptance of telemedicine in Germany and the possible reasons for its rejection would also be of interest. Finally, the causes of technical irregularities should be analyzed in detail to improve the long-term provision of telemedicine.

Practical Implications

In summary, our results suggest that the effective implementation of telemedical follow-up care ideally meets several conditions. First, the appropriateness of telemedicine should be individually assessed for each patient. In our study, age and sex did not significantly influence telemedicine choice. Nevertheless, physicians should consider whether the patient's condition and circumstances allow for a video consultation. Before implementing a video consultation in a clinic, criteria should be established to assist with patient selection. These criteria could be based on the inclusion and exclusion criteria of our study, complemented by clinic-dependent characteristics. In addition, suitable patients should be supported to overcome initial uncertainties.

Second, before each consultation, each patient should be assessed individually to determine whether a video consultation is sufficient or whether the patient should attend the clinic. Although physicians in our study would recommend a video consultation as the next appointment for most patients (73% in the intervention group and 64% in the control group), clinical consultations might still be necessary. Moreover, if any medical issues cannot be clarified in a video consultation or if problems occur, additional in-clinic treatment should always be possible, as was the case in this study. To be able to ensure patient safety in the long term, the monitoring, documentation, and control of adverse events is another indispensable factor in this context as well.

Our data refer to patients in orthopedic and trauma surgery. Nevertheless, our results and considerations for practical implications could be transferred to outpatient follow-up examinations in other specialties, such as general and visceral surgeries, if conversations and visual examinations are sufficient for the intended treatment. Therefore, our study could be used as a basis for decision-making regarding the use of telemedicine in different medical fields, supplemented by specialty-specific determinants.

Conclusions

Compared with international findings, this study highlights that telemedicine is an efficient option for patients in Germany with a broad range of indications in orthopedic and trauma surgery, especially for follow-up appointments. Most patients in the telemedicine group preferred their next follow-up appointment to be a video consultation rather than a standard in-clinic consultation. All patients in this study participated in telemedicine without any prior test run or support from staff, which corresponds to real-life conditions encountered in everyday clinical practice. Clearly, some consultations will always have to occur in hospitals, but telemedicine can be applied efficiently to a wide range of diagnoses and a wide range of patients, thus reducing the burden on patients, physicians, and clinical resources. The COVID-19 pandemic has acted as a catalyst for the widespread uptake of telemedicine. On the one hand, this provided a safe alternative to prevent infections. On the other hand, it demonstrated the benefits of telemedicine. This is why video consultations should find their way into health care beyond the COVID-19 pandemic as a supplement to clinical care.

Acknowledgments

The authors wish to thank Dr med Alexander Eicher for his support in organizing the study. This research did not receive any funding from an external body.

Conflicts of Interest

None declared.

Multimedia Appendix 1
International Classification of Diseases-10 codes of health conditions studied.
[DOCX File , 14 KB - jmir_v24i7e36996_app1.docx]

Multimedia Appendix 2
Box and whisker plots.
[DOCX File , 14988 KB - jmir_v24i7e36996_app2.docx]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 322 KB - jmir_v24i7e36996_app3.pdf]

References

1. Doty B, Zuckerman R, Finlayson S, Jenkins P, Rieb N, Heneghan S. How does degree of rurality impact the provision of surgical services at rural hospitals? J Rural Health 2008;24(3):306-310. [doi: 10.1111/j.1748-0361.2008.00173.x] [Medline: 18643809]
2. Schröder L, Flägel K, Goetz K, Steinhäuser J. Mobility concepts and access to health care in a rural district in Germany: a mixed methods approach. BMC Fam Pract 2018 May 02;19(1):47 [FREE Full text] [doi: 10.1186/s12875-018-0733-6] [Medline: 29720091]
3. van den Berg N, Schmidt S, Stentzel U, Mühlan H, Hoffmann W. Telemedizinische Versorgungskonzepte in der regionalen Versorgung ländlicher Gebiete : Möglichkeiten, Einschränkungen, Perspektiven. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2015 Apr;58(4-5):367-373. [doi: 10.1007/s00103-015-2134-5] [Medline: 25698122]
4. Federal Ministry of Health. The German healthcare system. Strong. Reliable. Proven. 2020. URL: https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/5_Publikationen/Gesundheit/Broschueren/200629_BMG_Das_deutsche_Gesundheitssystem_EN.pdf [accessed 2021-07-4]
5. WHO Europe. Coronavirus disease (COVID-19) pandemic. 2020. URL: https://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/novel-coronavirus-2019-ncov [accessed 2021-07-4]
6. Jansen T, Gathen M, Touet A, Goost H, Wirtz DC, Burger C, et al. Spine examination during COVID-19 pandemic via video consultation. Z Orthop Unfall 2021 Apr;159(2):193-201 [FREE Full text] [doi: 10.1055/a-1283-7160] [Medline: 33530112]
7. Jaenisch M, Kohlhof H, Touet A, Kehrer M, Cucchi D, Burger C, et al. Evaluation of the feasibility of a telemedical examination of the hip and pelvis - early lessons from the COVID-19 pandemic. Z Orthop Unfall 2021 Feb;159(1):39-46 [FREE Full text] [doi: 10.1055/a-1289-0779] [Medline: 33327028]
8. Abbas M, Robalo Nunes T, Martischang R, Zingg W, Iten A, Pittet D, et al. Nosocomial transmission and outbreaks of coronavirus disease 2019: the need to protect both patients and healthcare workers. Antimicrob Resist Infect Control 2021 Jan 06;10(1):7 [FREE Full text] [doi: 10.1186/s13756-020-00875-7] [Medline: 33407833]
9. Hirschmann MT, Hart A, Henckel J, Sadoghi P, Seil R, Mouton C. COVID-19 coronavirus: recommended personal protective equipment for the orthopaedic and trauma surgeon. Knee Surg Sports Traumatol Arthrosc 2020 Jun;28(6):1690-1698 [FREE Full text] [doi: 10.1007/s00167-020-06022-4] [Medline: 32342138]
10. Deutsches Ärzteblatt. Ärzte appellieren, Arztbesuche nicht aufzuschieben. 2020 Apr 29. URL: https://www.aerzteblatt.de/nachrichten/112413/Aerzte-appellieren-Arztbesuche- [accessed 2021-07-5]
11. Rubin R. COVID-19's crushing effects on medical practices, some of which might not survive. JAMA 2020 Jul 28;324(4):321-323. [doi: 10.1001/jama.2020.11254] [Medline: 32556122]
12. Bundesministerium für Gesundheit. Coronavirus-Pandemie (SARS-CoV-2): Chronik bisheriger Maßnahmen und Ereignisse. 2020. URL: https://www.bundesgesundheitsministerium.de/coronavirus/chronik-coronavirus.html [accessed 2021-07-5]
13. Kumar S, Kumar A, Kumar M, Kumar A, Arora R, Sehrawat R. Feasibility of telemedicine in maintaining follow-up of orthopaedic patients and their satisfaction: a preliminary study. J Clin Orthop Trauma 2020 Oct;11(Suppl 5):S704-S710 [FREE Full text] [doi: 10.1016/j.jcot.2020.07.026] [Medline: 32837105]
14. Shachar C, Engel J, Elwyn G. Implications for telehealth in a postpandemic future: regulatory and privacy issues. JAMA 2020 Jun 16;323(23):2375-2376. [doi: 10.1001/jama.2020.7943] [Medline: 32421170]
15. World Health Organization. A Health Telematics Policy. 1998. URL: https://apps.who.int/iris/bitstream/handle/10665/63857/WHO_DGO_98.1.pdf?sequence=1&isAllowed=y [accessed 2021-07-5]
16. Parisien RL, Shin M, Constant M, Saltzman BM, Li X, Levine WN, et al. Telehealth utilization in response to the novel coronavirus (COVID-19) pandemic in orthopaedic surgery. J Am Acad Orthop Surg 2020 Jun 01;28(11):e487-e492 [FREE Full text] [doi: 10.5435/JAAOS-D-20-00339] [Medline: 32459409]
17. 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 [FREE Full text] [doi: 10.15585/mmwr.mm6943a3] [Medline: 33119561]
18. 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. J Am Acad Orthop Surg Glob Res Rev 2020 Sep;4(9):e20.00127-e20.00113 [FREE Full text] [doi: 10.5435/JAAOSGlobal-D-20-00127] [Medline: 33939394]
19. Williams AM, Bhatti UF, Alam HB, Nikolian VC. The role of telemedicine in postoperative care. Mhealth 2018 May 2;4:11 [FREE Full text] [doi: 10.21037/mhealth.2018.04.03] [Medline: 29963556]
20. Petersen W, Karpinski K, Backhaus L, Bierke S, Häner M. A systematic review about telemedicine in orthopedics. Arch Orthop Trauma Surg 2021 Oct;141(10):1731-1739 [FREE Full text] [doi: 10.1007/s00402-021-03788-1] [Medline: 33635399]
21. 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]
22. 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 Jan 04;24(1):e31752 [FREE Full text] [doi: 10.2196/31752] [Medline: 34854815]

23. Buvik A, Bugge E, Knutsen G, Småbrekke A, Wilsgaard T. Quality of care for remote orthopaedic consultations using telemedicine: a randomised controlled trial. BMC Health Serv Res 2016 Sep 08;16(1):483 [FREE Full text] [doi: 10.1186/s12913-016-1717-7] [Medline: 27608768]
24. Harle D, Ilyas S, Darrah C, Tucker K, Donell S. Community-based orthopaedic follow-up. Is it what doctors and patients want? Ann R Coll Surg Engl 2009 Jan;91(1):66-70 [FREE Full text] [doi: 10.1308/003588409X359105] [Medline: 19126337]
25. Sharareh B, Schwarzkopf R. Effectiveness of telemedical applications in postoperative follow-up after total joint arthroplasty. J Arthroplasty 2014 May;29(5):918-22.e1. [doi: 10.1016/j.arth.2013.09.019] [Medline: 24342278]
26. Statistisches Bundesamt (Destatis). Gesundheit - Grunddaten der Krankenhäuser - Fachserie 12 Reihe 6.1.1. 2019. URL: https://tinyurl.com/2p8dsdk4 [accessed 2021-07-5]
27. Buvik A, Bergmo TS, Bugge E, Smaabrekke A, Wilsgaard T, Olsen JA. Cost-effectiveness of telemedicine in remote orthopedic consultations: randomized controlled trial. J Med Internet Res 2019 Feb 19;21(2):e11330 [FREE Full text] [doi: 10.2196/11330] [Medline: 30777845]
28. Buvik A, Bugge E, Knutsen G, Småbrekke A, Wilsgaard T. Patient reported outcomes with remote orthopaedic consultations by telemedicine: a randomised controlled trial. J Telemed Telecare 2019 Sep;25(8):451-459. [doi: 10.1177/1357633X18783921] [Medline: 29973130]
29. Sathiyakumar V, Apfeld JC, Obremskey WT, Thakore RV, Sethi MK. Prospective randomized controlled trial using telemedicine for follow-ups in an orthopedic trauma population: a pilot study. J Orthop Trauma 2015 Mar;29(3):e139-e145. [doi: 10.1097/BOT.0000000000000189] [Medline: 24983434]
30. Kane LT, Thakar O, Jamgochian G, Lazarus MD, Abboud JA, Namdari S, et al. The role of telehealth as a platform for postoperative visits following rotator cuff repair: a prospective, randomized controlled trial. J Shoulder Elbow Surg 2020 Apr;29(4):775-783. [doi: 10.1016/j.jse.2019.12.004] [Medline: 32197766]
31. Holderried M, Schlipf M, Höper A, Meier R, Stöckle U, Kraus TM. Chancen und Risiken der Telemedizin in der Orthopädie und Unfallchirurgie. Z Orthop Unfall 2018 Feb;156(1):68-77. [doi: 10.1055/s-0043-116941] [Medline: 28837980]
32. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c869 [FREE Full text] [doi: 10.1136/bmj.c869] [Medline: 20332511]
33. Kassenärztliche Bundesvereinigung. Zufriedenheit in der ambulanten Versorgung - Qualität aus Patientenperspektive. 2019. URL: https://www.kbv.de/media/sp/ZAP_Fragebogen_Deutsch.pdf [accessed 2020-02-11]
34. Kassenärztliche Bundesvereinigung. Zufriedenheit in der ambulanten Versorgung - Qualität aus Patientenperspektive (Englisch). 2019. URL: https://www.kbv.de/media/sp/ZAP_Fragebogen_Englisch.pdf [accessed 2022-03-18]
35. 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]
36. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
37. Haukipuro K, Ohinmaa A, Winblad I, Linden T, Vuolio S. The feasibility of telemedicine for orthopaedic outpatient clinics--a randomized controlled trial. J Telemed Telecare 2000;6(4):193-198. [doi: 10.1258/1357633001935347] [Medline: 11027118]
38. Marsh JD, Bryant DM, MacDonald SJ, Naudie DD, McCalden RW, Howard JL, et al. Feasibility, effectiveness and costs associated with a Web-based follow-up assessment following total joint arthroplasty. J Arthroplasty 2014 Sep;29(9):1723-1728. [doi: 10.1016/j.arth.2014.04.003] [Medline: 24881023]
39. Chae S, Lee YJ, Han HR. Sources of health information, technology access, and use among non-english-speaking immigrant women: descriptive correlational study. J Med Internet Res 2021 Oct 29;23(10):e29155 [FREE Full text] [doi: 10.2196/29155] [Medline: 34714249]
40. Shaw J, Brewer LC, Veinot T. Recommendations for health equity and virtual care arising from the COVID-19 pandemic: narrative review. JMIR Form Res 2021 Apr 05;5(4):e23233 [FREE Full text] [doi: 10.2196/23233] [Medline: 33739931]
41. Braekman E, Berete F, Charafeddine R, Demarest S, Drieskens S, Gisle L, et al. Measurement agreement of the self-administered questionnaire of the Belgian Health Interview Survey: paper-and-pencil versus Web-based mode. PLoS One 2018 May 21;13(5):e0197434 [FREE Full text] [doi: 10.1371/journal.pone.0197434] [Medline: 29782504]
42. Braekman E, Charafeddine R, Demarest S, Drieskens S, Berete F, Gisle L, et al. Comparing Web-based versus face-to-face and paper-and-pencil questionnaire data collected through two Belgian health surveys. Int J Public Health 2020 Jan;65(1):5-16. [doi: 10.1007/s00038-019-01327-9] [Medline: 31993674]
43. Marsh J, Bryant D, MacDonald SJ, Naudie D, Remtulla A, McCalden R, et al. Are patients satisfied with a Web-based followup after total joint arthroplasty? Clin Orthop Relat Res 2014 Jun;472(6):1972-1981 [FREE Full text] [doi: 10.1007/s11999-014-3514-0] [Medline: 24562873]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials

RCT: randomized controlled trial
VAS: visual analog scale

Edited by G Eysenbach; submitted 02.02.22; peer-reviewed by E van der Velde, B Wolfe-Piggott, K Lam; comments to author 24.02.22; revised version received 31.05.22; accepted 12.06.22; published 27.07.22.

Please cite as:
Muschol J, Heinrich M, Heiss C, Knapp G, Repp H, Schneider H, Thormann U, Uhlar J, Unzeitig K, Gissel C
Assessing Telemedicine Efficiency in Follow-up Care With Video Consultations for Patients in Orthopedic and Trauma Surgery in Germany: Randomized Controlled Trial
J Med Internet Res 2022;24(7):e36996
URL: https://www.jmir.org/2022/7/e36996
doi:10.2196/36996
PMID:35896015

Original Paper

Effects of an Explicit Value Clarification Method With Computer-Tailored Advice on the Effectiveness of a Web-Based Smoking Cessation Decision Aid: Findings From a Randomized Controlled Trial

Thomas Gültzow[1,2], PhD; Eline Suzanne Smit[3], PhD; Rik Crutzen[1], Prof Dr; Shahab Jolani[4], PhD; Ciska Hoving[1], PhD; Carmen D Dirksen[5], Prof Dr

[1]Department of Health Promotion, Care and Public Health Research Institute, Maastricht University, Maastricht, Netherlands

[2]Department of Work & Social Psychology, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, Netherlands

[3]Department of Communication Science, Amsterdam School of Communication Research, University of Amsterdam, Amsterdam, Netherlands

[4]Department of Methodology and Statistics, Care and Public Health Research Institute, Maastricht University, Maastricht, Netherlands

[5]Department of Clinical Epidemiology and Medical Technology Assessment, Care and Public Health Research Institute, Maastricht University Medical Centre, Maastricht, Netherlands

Corresponding Author:
Thomas Gültzow, PhD
Department of Work & Social Psychology
Faculty of Psychology and Neuroscience
Maastricht University
PO Box 616
Maastricht, 6200 MD
Netherlands
Phone: 31 433881511
Email: thomas.gultzow@maastrichtuniversity.nl

Abstract

Background: Smoking continues to be a driver of mortality. Various forms of evidence-based cessation assistance exist; however, their use is limited. The choice between them may also induce decisional conflict. Offering decision aids (DAs) may be beneficial; however, insights into their effective elements are lacking.

Objective: This study tested the added value of an effective element (ie, an "explicit value clarification method" paired with computer-tailored advice indicating the most fitting cessation assistance) of a web-based smoking cessation DA.

Methods: A web-based randomized controlled trial was conducted among smokers motivated to stop smoking within 6 months. The intervention group received a DA with the aforementioned elements, and the control group received the same DA without these elements. The primary outcome measure was 7-day point prevalence abstinence 6 months after baseline (time point 3 [t=3]). Secondary outcome measures were 7-day point prevalence of abstinence 1 month after baseline (time point 2 [t=2]), evidence-based cessation assistance use (t=2 and t=3), and decisional conflict (immediately after DA; time point 1). Logistic and linear regression analyses were performed to assess the outcomes. Analyses were conducted following 2 (decisional conflict) and 3 (smoking cessation) outcome scenarios: complete cases, worst-case scenario (assuming that dropouts still smoked), and multiple imputations. A priori sample size calculation indicated that 796 participants were needed. The participants were mainly recruited on the web (eg, social media). All the data were self-reported.

Results: Overall, 2375 participants were randomized (intervention n=1164, 49.01%), of whom 599 (25.22%; intervention n=275, 45.91%) completed the DAs, and 276 (11.62%; intervention n=143, 51.81%), 97 (4.08%; intervention n=54, 55.67%), and 103 (4.34%; intervention n=56, 54.37%) completed time point 1, t=2, and t=3, respectively. More participants stopped smoking in the intervention group (23/63, 37%) than in the control group (14/52, 27%) after 6 months; however, this was only statistically significant in the worst-case scenario (crude P=.02; adjusted P=.04). Effects on the secondary outcomes were only observed for smoking abstinence after 1 month (15/55, 27%, compared with 7/46, 15%, in the crude and adjusted models, respectively; P=.02) and for cessation assistance uptake after 1 month (26/56, 46% compared with 18/47, 38% only in the crude model; P=.04) and

6 months (38/61, 62% compared with 26/50, 52%; crude P=.01; adjusted P=.02) but only in the worst-case scenario. Nonuse attrition was 34.19% higher in the intervention group than in the control group (P<.001).

Conclusions: Currently, we cannot confidently recommend the inclusion of explicit value clarification methods and computer-tailored advice. However, they might result in higher nonuse attrition rates, thereby limiting their potential. As a lack of statistical power may have influenced the outcomes, we recommend replicating this study with some adaptations based on the lessons learned.

Trial Registration: Netherlands Trial Register NL8270; https://www.trialregister.nl/trial/8270

International Registered Report Identifier (IRRID): RR2-10.2196/21772

(J Med Internet Res 2022;24(7):e34246) doi:10.2196/34246

KEYWORDS

digital health; decision-making; decision support technique; decision aid; smoking; smoking cessation; informed decision-making; decision support; decision support tool; eHealth; evidence-based medicine; value clarification method

Introduction

Background

Smoking continues to be a major driver of global mortality [1], a trend that is mirrored in the Netherlands [2,3], which showcases that smoking prevention is imperative. However, as 21.7% of the adult Dutch population still smokes, according to recent figures [2], it is also of the utmost importance to invest in effective and evidence-based cessation assistance. Several cessation assistance tools are currently recognized as effective in aiding individuals in successfully achieving smoking abstinence. These range from behavioral interventions (eg, counseling [4]) to nicotine replacement therapy [5] and prescription medication [6]. Unfortunately, evidence-based cessation assistance is underused in various countries, including the Netherlands [7].

A common barrier to cessation assistance use is incorrect knowledge of the safety and efficacy of cessation assistance [8]. Therefore, increasing individuals' knowledge may be a worthwhile avenue to explore when it comes to successfully increasing individuals' cessation rates. In particular, being uninformed can lead to a state of uncertainty about which course of action to take—better known as decisional conflict [9]—which, in turn, is known to increase the chances of decision delay [9,10]. Therefore, individuals who decide to stop smoking but are uninformed about the possibilities of cessation assistance might delay the decision on *how* to stop smoking or simply decide to use the most common approach: attempting to quit smoking without the use of cessation assistance [11]. That said, knowledge provision alone is often insufficient to facilitate behavior change. This is commonly referred to as the knowledge-behavior gap in health promotion research [12,13]. Thus, knowledge provision alone might not be enough to support smokers motivated to quit smoking in their decisions about smoking cessation assistance. However, providing individuals who are motivated to stop smoking with decision support that also includes accurate information (next to other decision support elements) about cessation assistance tools might decrease decisional conflict. This, in turn, might facilitate smoking cessation efforts and ultimately increase the chances of long-term smoking abstinence.

Such decision support can be provided in the form of decision aids (DAs), which are interventions specifically designed to facilitate decisional processes [14]. A recent systematic review by Moyo et al [15] showed that DAs can be beneficial for smoking cessation, although traditionally, DAs have most often been applied to treatment and screening decisions rather than lifestyle-related decisions [14]. For example, BinDhim et al [16] showed that smoking cessation DAs can result in an increase in continuous abstinence at 1, 3, and 6 months compared with an intervention containing information only. Participants in the DA group were also more likely to have made an informed choice and showed fewer decisional conflict. However, information about the effective elements of smoking cessation DAs is currently lacking, and the only smoking cessation DA that has been previously studied in a Dutch context [17] has shown several limitations: it was largely paper based, thereby limiting widespread dissemination; it lacked an interactive design, although interactivity has been shown to positively influence factors such as information comprehensibility and attitudinal beliefs [18,19]; and it did not explicitly include methods of helping end users become aware of what is important to them personally (in the DA literature, this is often referred to as value clarification [20]), although this is regarded as an active DA element [21,22]. Moreover, interestingly, this DA had a positive effect on smoking cessation success but not on the uptake of cessation assistance [17]. Improving cessation assistance uptake might further increase the effectiveness of smoking cessation DAs, and overcoming the aforementioned limitations could play a promising role in achieving this.

To illustrate, given that a lack of knowledge is considered a barrier to cessation assistance use, adding interactive elements to a smoking cessation DA might be particularly helpful as interactive elements can improve information comprehensibility and positively influence individuals' beliefs [18,19]. In addition to the use of interactive elements, explicitly devoting attention to smokers' personal values could also positively influence the effects of smoking cessation DAs. The International Patient Decision Aid Standards Collaboration (IPDAS) regards these so-called value clarification methods (VCMs) as active DA elements [21,22]. VCMs can be divided into 2 different formats. *Explicit* VCMs refer to exercises that actively engage users in an activity to clarify what is important to them personally (eg,

scoring certain statements), whereas *implicit* VCMs refer to the provision of static information that is specifically linked to the decision at hand. In other words, explicit VCMs include an element of interactivity that implicit VCMs lack. Recently, scholars have started to pay more attention to studying the added value of explicit (as opposed to implicit) VCMs. Previous studies have shown that explicit VCMs seem to be more effective than implicit VCMs in terms of decision-making processes [23], especially in the long run [24] and when people are supported in understanding the implications of their clarified values [25,26]. An approach to facilitate the understanding of the implications of clarified values is to show participants the options that best fit their clarified values [25]; for example, by providing computer-tailored advice based on answers provided in the explicit VCM. However, to date, it has not been studied whether the addition of explicit VCMs paired with such advice positively affects smoking cessation outcomes. To advance our understanding of the effectiveness of smoking cessation DAs and support more people in the Netherlands to quit smoking successfully, we developed a web-based smoking cessation DA (called *VISOR*) that includes interactivity and an explicit VCM paired with computer-tailored advice and studied its effects in a randomized controlled trial (RCT).

The Smoking Cessation DA VISOR

In accordance with the IPDAS guidelines for DA development [27,28], *VISOR* was developed by a steering team (TG, ESS, CDD, and CH) that led a development process involving both professional experts and potential end users, for example, by assessing their needs and opinions before the initial development [29] and by conducting usability tests—this development process is described in detail elsewhere [28].

Moreover, we used the self-determination theory (SDT) [30] as the theoretical background. The SDT revolves around the formation of motivation and posits that 3 psychological needs (ie, the needs for autonomy, relatedness, and competence) are essential to developing autonomous motivation (ie, motivation that emanates from oneself and is not primarily externally motivated) [30,31]. Autonomous motivation is assumed to have a greater influence on long-term behavior change compared with more controlled forms of motivation [32]. The SDT is particularly well-suited to support DA developers because it focuses on perceived autonomy as opposed to theories that are often used to develop more persuasive interventions. DAs are similarly geared toward personal autonomy. Therefore, the SDT supported us in developing a DA that not only provided accurate information but also helped motivate end users, for example, by framing the information in the DA autonomy supportively. Finally, *VISOR* took a stepwise approach to facilitate autonomous decision-making, which is concordant with personal values and smoking behavior. *VISOR* comprised 8 sections that are described in Textbox 1.

VISOR was a stand-alone, 1-time intervention meant to support adult smokers in the general population in their decision to use smoking cessation assistance and did not have to be used together with a health care professional. That said, *VISOR* could be used to prepare patients and clients for a health care consultation about smoking (cessation), and certain cessation assistance options (eg, prescription medication) required a health care provider to prescribe them. Therefore, if *VISOR* users chose to use a cessation assistance option that required a prescription from a health care provider, they were advised to contact their health care provider to gain access to this specific option. More in-depth information on *VISOR*, including the specific theoretical underpinnings of each step, can be found elsewhere [28]. Examples of the information section and explicit VCMs are shown in Figure 1 (information section [33]) and Figure 2 (explicit VCMs). The screenshots were translated from the original Dutch to English to facilitate understanding.

Textbox 1. Overview of VISOR's sections.

Overview of VISOR's sections

1. Information section explaining the decision at hand, as well as all the cessation assistance available in the Netherlands
2. Optional knowledge quiz
3. Brief smoking assessment
4. Intuitive decision between different clusters of cessation assistance tools:
   - Behavioral support
   - Nicotine replacement therapy (NRT)
   - Combination of behavioral support, NRT, and prescription medication
   - Combination of behavioral support and NRT
   - Combination of behavioral support and prescription medication
   - Other (non–evidence-based) cessation assistance
   - No cessation assistance at all
5. Intermediate advice to use a combination of behavioral and pharmacological cessation assistance tools, for users who chose the following:
   - Behavioral support only, or NRT only, while also indicating that they smoke >10 cigarettes on a normal day and/or have made ≥1 smoking cessation attempt or attempts in the past (in step 3)
   - Non–evidence-based cessation assistance or no cessation assistance at all in step 4 regardless of their answers in step 3
6. Explicit value clarification method (VCM) for users who chose evidence-based cessation assistance tools in steps 4 or 5, where users were asked to rate certain statements regarding cessation assistance characteristics (eg, "I prefer a stop method that works better, even if that means that I have to leave the house").
   - Users only rated statements for options that belonged to the cluster of cessation assistance options they selected in the previous step or steps
7. Computer-tailored advice based on the explicit VCM, including an optional ranking of all options; only when it was possible to give clear advice, (ie, if users' scores did not suggest that >2 cessation assistance tools were equally suitable based on their indicated values)
8. Access information on how to obtain the chosen cessation assistance (eg, nicotine patches)

Figure 1. Screenshot of the information section in the decision aid (original text translated from Dutch); the displayed icon array has been created using IconArray [33].

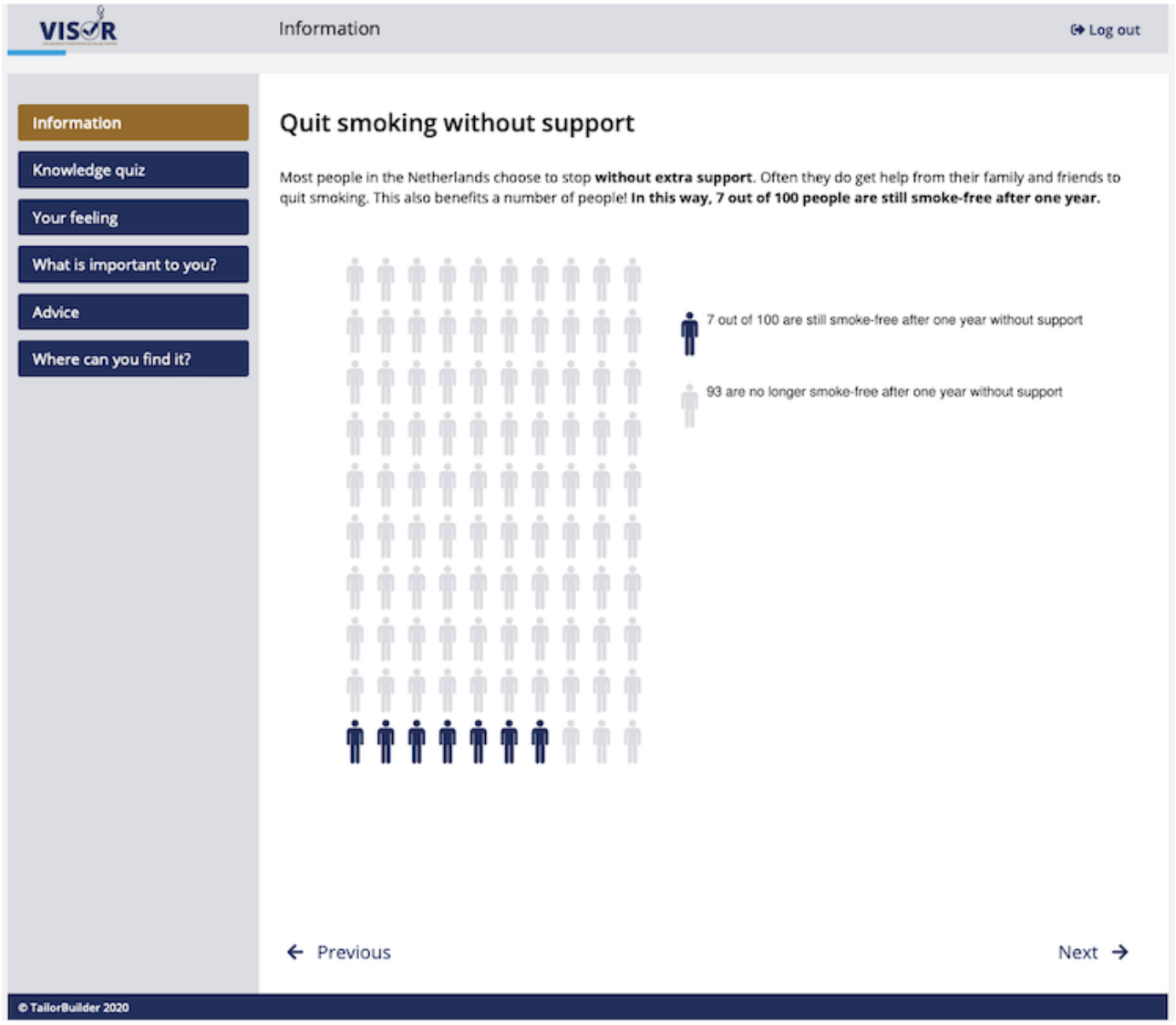
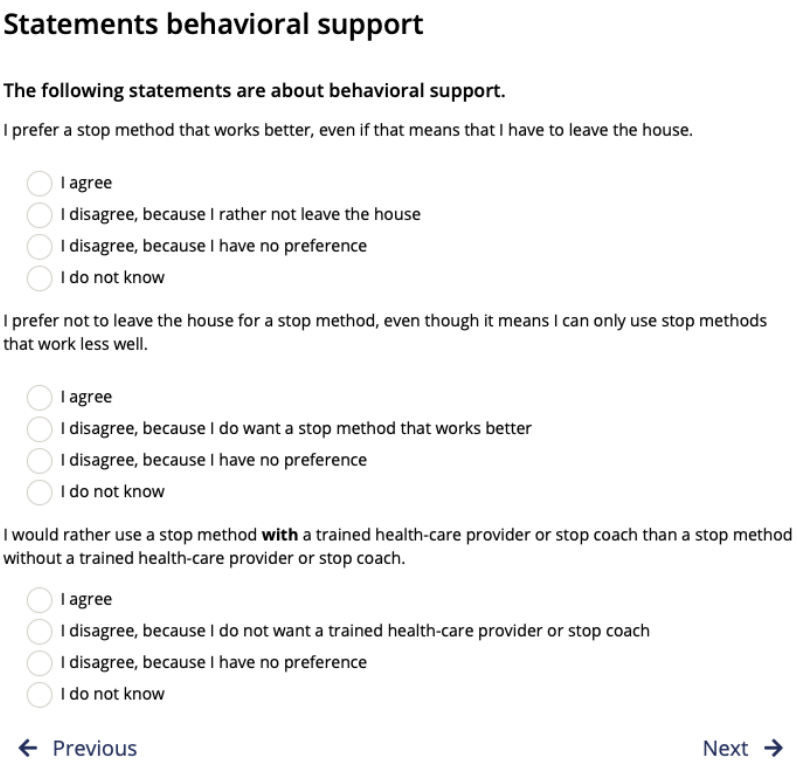


Figure 2. Screenshot of a part of the VCM in the decision aid (original text translated from Dutch). VCM: value clarification method.



Study Goals and Hypotheses

In this study, the behavioral (ie, smoking abstinence and cessation assistance use) and decisional effects (ie, decisional conflict) of an explicit VCM paired with computer-tailored advice (within a smoking cessation DA) are reported. An RCT was conducted to investigate these effects. Specifically, we tested the following hypotheses (as described in the study protocol [28] and the Netherlands Trial Register [NL8270]):

- Hypothesis 1a or 1b is that a DA with explicit VCM and computer-tailored advice will lead to a statistically significant increase in smoking abstinence after 1 month (hypothesis 1a) and 6 months (hypothesis 1b) compared with a DA without explicit VCM and computer-tailored advice.
- Hypothesis 2a or 2b is that a DA with explicit VCM and computer-tailored advice will lead to a statistically significant increase in evidence-based cessation assistance use after 1 month (hypothesis 2a) and 6 months (hypothesis 2b) compared with a DA without explicit VCM and computer-tailored advice.
- Hypothesis 3 is that a DA with explicit VCM and computer-tailored advice will lead to a statistically significant decrease in decisional conflict (state of uncertainty about which course of action to take) immediately after using the DA compared with a DA without an explicit VCM and computer-tailored advice.

Methods

Overview

An RCT was conducted in line with the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist [34]. However, we deviated from the checklist in one aspect: participants were only invited for follow-up measurements if they completed one of the DAs; that is, if they completed the intervention until the end. This was because we wanted to ensure that the participants we included in the analysis received additional intervention elements as those were the focus of our RCT.

Ethics Approval

The study did not fall under the scope of the Medical Research Involving Human Subjects Act, as indicated by the Medical Ethics Committee of Zuyderland, the Netherlands (16-N-227; International Registered Report Identifier: RR2-10.2196/21772), and the development of *VISOR* and the accompanying studies (eg, the aforementioned needs assessment [29]) were funded by the Dutch Cancer Society (UM2015-7744). Study materials (including the underlying data and statistical scripts) can be found on the open science framework (OSF) website [35].

Primary and Secondary Outcomes

The primary outcome of this study was 7-day point prevalence after 6 months (corresponding to hypothesis 1b). Secondary outcomes were 7-day point prevalence after 1 month (corresponding to hypothesis 1a), evidence-based cessation assistance use after 1 and 6 months (corresponding to hypothesis

2a or 2b), and decisional conflict directly after using the DA (corresponding to hypothesis 3).

Sample Size

An a priori power calculation was conducted based on the primary outcome measure of 7-day point prevalence abstinence and the only other RCT in the Netherlands, which tested the effect of a smoking cessation DA in which a significant effect (20.2% vs 13.6%) was found at 6 months [17]. To be able to significantly (α=.05; β=.20) detect the same effect in a 1-sided test, 398 smokers per arm were necessary at the end of the trial (796 in total). Considering 50% attrition over the study period, we aimed to include 1592 smokers at baseline.

Study Population

Participants were included if they were (1) currently smoking, (2) motivated to stop smoking within 6 months, (3) aged between 18 and 100 years, (4) able to understand Dutch, and (5) had access to the internet and the necessary internet literacy (skills) to use the DA. The last 2 inclusion criteria were not actively screened but were deemed inherent to participation. Participants were excluded if they did not meet the inclusion criteria or exclusively used e-cigarettes. As described in the study protocol [28], participants were mainly recruited on the web to reflect the web-based nature of *VISOR*, and the entire trial was web-based (ie, there were no offline contacts). Recruitment took place mainly by using paid social media advertisements and unpaid social media posts on project accounts, [36] which were also shared on the team members' accounts, their respective institutions, and other relevant organizations within the Netherlands. In addition, *VISOR* was featured in regional media (eg, a newspaper interview), and we used a project website with a direct access point to *VISOR* via a clickable button. Owing to the small influx of participants (especially after the start of the COVID-19 pandemic), we decided to deploy other additional recruitment activities as well, such as a study call, information in a relevant Dutch journal for general practitioners [37], and student pools at the Universities of Amsterdam and Maastricht. All (potential) participants received information about the content of *VISOR*, the duration of the study (including the number of follow-up measurements), and the compensation that they received for completing the last measurement (€10 [US $12.17]). Throughout the study, participants also received information on the duration of *VISOR* and the questionnaires. The participants received no information regarding the differences between the intervention and control groups to avoid bias in the results of the trial. The students at the University of Amsterdam received research credits instead of monetary compensation. All recruitment materials (eg, the project website) included a display of the project team's institutional affiliations in some form.

Intervention and Comparator Groups

Participants in the intervention group received the DA as described in the *Introduction* section (see *The Smoking Cessation DA VISOR* section), whereas participants in the control group received the same DA, excluding the explicit VCM and computer-tailored advice; that is, steps 6 and 7 described in Textbox 1 were skipped. The only other (small) difference was that participants in the intervention group were immediately directed toward the end after they had chosen to not use evidence-based cessation assistance (ie, step 8—access information—was skipped), which was not the case for the control group. Thus, both groups had a chance to re-evaluate their choice, as the intervention group was offered a chance to re-evaluate their choice during the additional elements. Neither the DA received by the intervention group nor that received by the control group changed throughout the trial. Additional information can be found in the study protocol [28].

Trial Flow and Measurement Instruments

Overview

In total, the study comprised 4 fully automated and web-based contact moments: time point 0 for the baseline questionnaire and *VISOR*, time point 1 (t=1) directly after participants had used *VISOR*, time point 2 (t=2) after 1 month, and time point 3 (t=3) after 6 months. Participants were asked to fill in each follow-up questionnaire if they made use of the entire DA, even if they did not fill in one of the other follow-up questionnaires. To avoid high attrition rates, participants received either 1 automatic reminder after a week (if they had not filled in a follow-up questionnaire at all) or 2 after 2 days and a week (if they had already started filling in at least part of a follow-up questionnaire). Participants who started using *VISOR* or started filling in the baseline questionnaire (time point 0) without finishing it also received 2 automatic reminders (after 2 days and a week). In the last reminder for t=3, participants were also offered the option to share only their answers regarding the primary outcome (ie, 7-day point prevalence abstinence) with the research team. All data were self-assessed. If available, we used previously validated measurements [9] and measurements that were previously used in a Dutch context [38]; if possible, we used measurements that were previously used in self-administered web-based studies [39,40]. For more information, refer to the Checklist for Reporting Results of Internet E-Surveys checklist in Multimedia Appendix 1. In the case of psychological constructs collected using multiple items (eg, decisional conflict), we assessed scale quality, as proposed by Crutzen and Peters [41], using the *Rosetta Stats* package in R (R Foundation for Statistical Computing) [42,43] in two steps: (1) investigating scale structure by exploratory factor analysis and (2) calculating omega (Ω) [44] as a less biased alternative to (Cronbach) α.

Participants were registered for the study via a web-based form, which included their provision of informed consent and the creation of an account. Before account creation, participants were automatically randomized into either the intervention or the control group by the web-based platform on which questionnaires and *VISOR* were hosted, allocating approximately 50% of the respondents to either group. The end users were blinded to their allocated groups. Immediately after registration, the participants were asked to complete the baseline questionnaire. A visual representation of the trial flow can be found in the study protocol [28].

Baseline Measurements: Directly Before the DA

Demographic information was collected based on 3 criteria: age, gender, and education. Smoking behavior was collected regarding the used tobacco products, amount of tobacco consumption per product per day, past cessation attempts, amount of past cessation attempts (for people who previously attempted to stop smoking), and cessation assistance use in the past 6 months. If the participants indicated that they had used cessation assistance, they were also asked what had been used. Additional information regarding demographic information and smoking behavior can be found in Multimedia Appendix 2 [45-47].

Nicotine dependence was measured using the Revised Fagerström Test for Nicotine Dependence (FTND-R), which has shown better psychometric properties than the unrevised version [48] and was verified to be unidimensional within our sample (Ω=0.76; further information can be found on the OSF [35]). We changed the wording of the items slightly (eg, *smoking moment* instead of *cigarette*) to include other tobacco products as well and left out the item relating to the amount of cigarette consumption. Rather, we created a composite score based on participants' answers to the item about the number of tobacco products, expressed as the number of cigarettes, with 1 hand-rolled cigarette and 1 other or a cannabis product equaling 1 cigarette, 1 pipe equaling 2.5 cigarettes, and 1 cigar equaling 4 cigarettes [38,49]. Subsequently, this composite score was recoded in line with the FTND-R (0=0-10 cigarettes, 1=11-20 cigarettes, 2=21-30 cigarettes, and 3=more than 30 cigarettes). We had to exclude e-cigarette use for this composite score as it was unclear how the answer categories of the e-cigarette use item related to the amount of cigarette consumption. Subsequently, a composite score was created by summing this item with the 5 other FTND-R items, resulting in 1 FTND-R score per individual ranging from 0 to 16, with 0 indicating no dependence.

Finally, the stage of decision-making was measured with 1 item: "Have you thought about how to quit smoking at this point? Choose the answer that best suits your situation; 1=I haven't begun to think about the choices, 2=I haven't begun to think about the choices, but am interested in doing so, 3=I am considering the options now, 4=I am close to selecting an option, 5=I have already made a decision, but am still willing to reconsider, 6=I have already made a decision and am unlikely to change my mind" [50].

Follow-up at t=1: Directly After the DA

After the DA, we measured decisional conflict (secondary outcome) using the Decisional Conflict Scale [9], which was also verified to be unidimensional in our sample (Ω=0.98; further information can be found on the OSF [35]). We used all 16 items using the original statement format with 5 response categories (0=strongly agree, 1=agree, 2=neither agree nor disagree, 3=disagree, and 4=strongly disagree) and created a composite score as described in the user manual provided by O'Connor [51]: individuals' scores were summed, divided by 16, and multiplied by 25. Thus, every participant had a score ranging from 0 (no decisional conflict) to 100 (extremely high decisional conflict). Scores >37.5 were generally regarded as being associated with decision delay or being unsure about decision implementation [51].

Follow-ups at t=2 and t=3: 1 Month and 6 Months After Baseline

After 1 month and 6 months, participants were queried regarding the choice they made (ie, the implemented decision; secondary outcome) and whether they were able to abstain from smoking in the previous 7 days (ie, 7-day point prevalence abstinence; secondary and primary outcomes).

Data Analysis

All analyses were performed in R [43] using the integrated development environment of RStudio [52]. First, descriptive analyses were conducted to assess the sample characteristics.

Second, to determine which factors influenced nonuse attrition (ie, attrition during the intervention) and dropout attrition (ie, not returning to the follow-ups) [53], we first compared participants at baseline who did not finish *VISOR* with those who finished *VISOR* using the Wilcoxon rank-sum test, Mood median test, chi-square test, or Fisher exact test, depending on the variable in question. Subsequently, we checked the results of these univariate analyses by using logistic regression with nonuse attrition as the outcome and all significant variables from the univariate analyses as predictors. Regarding dropout attrition, we compared participants at t=1, t=2, and t=3 with those lost to follow-up (after having used the intervention) using the same approach as described previously.

Third, logistic regression was used to test hypotheses 1a, 1b, 2a, and 2b. First, we conducted crude analyses in which we only included the allocated group (ie, the intervention group compared with the control group) as a predictor and 7-day point prevalence abstinence as the outcome, followed by fully adjusted analyses in which we corrected for age, gender, education, and the FTND-R [28,54]. All covariates were selected a priori [28,54,55]; however, because of multicollinearity issues in some of the adjusted models testing hypotheses 2a and 2b, we needed to recode the educational variable into low and high rather than low, medium, and high (as for the other analyses). This was only done for the models where this was the case. In addition, because of the size of the 2 very small gender (identity) groups (ie, nonbinary participants and participants who preferred not to state their gender), we were unable to include participants belonging to these groups in all adjusted models. Therefore, we decided to include the data of those participants in the crude models but not in the adjusted models. To test the robustness of the results, logistic regression was conducted according to three different approaches: (1) complete cases only, (2) worst-case scenario (dropout respondents were considered to still smoke; ie, penalized imputation), and (3) multiple imputations (MIs) using the *mice* package [56]. Variables that were included in the imputation model can be found in Multimedia Appendix 3. In the MI models, we accounted for a technical mistake described in the *Changes From the Study Protocol* section. As we had directional hypotheses, we conducted 1-sided tests as planned a priori [28]. To calculate the P values, we used the following formulas: (1) $P/2$ if the effect moved in the hypothesized direction and (2) $1 – (P/2)$ if

the effect moved in the other direction. Similarly, we swapped the upper bound of the calculated CIs to infinity. This was only done for the intervention effect as we had no specific directional hypotheses regarding the covariates.

Finally, hypothesis 3 was tested using linear regression. Again, we started with the crude analyses, using the allocated group as a predictor and the Decisional Conflict Scale as the outcome, followed by fully adjusted analyses corrected for covariates selected a priori (ie, age, gender, education, and stage of decision-making) [28,54,55]. Again, we had to exclude participants from certain analyses because they belonged to a very small gender (identity) group. Regarding hypothesis 3, we used only (1) complete cases and (2) MI using the *mice* package [56]. Variables that were included in the imputation model can be found in Multimedia Appendix 4. Again, we calculated the values for the 1-sided tests as described previously; however, as we expected a negative effect, we swapped the lower bound of the calculated CIs with infinity. In complete case analyses, all participants who filled in the respective outcome measures (eg, everyone who filled in the entire Decisional Conflict Scale) were included, even when participants did not finish the entire follow-up questionnaire.

Changes From the Study Protocol

In the original study protocol, and as described in the Netherlands Trial Register, participants were contacted 4 times after having used *VISOR* (ie, directly after the DA and after 1, 6, and 12 months). Unfortunately, we had to extend our recruitment period because of the COVID-19 pandemic, and consequently, the recruitment period lasted for approximately 12 months (ie, 6 months longer than initially planned). Consequently, given the maximum project duration funded by the Dutch Cancer Society, we did not perform the 12-month follow-up measurements. Therefore, the original primary outcome (ie, 7-day point prevalence after 12 months) had to be adjusted, and the 7-day point prevalence after 6 months was ultimately used as the primary outcome. This change was communicated to and approved by the Dutch Cancer Society. Other changes were not communicated and approved by them as this was not deemed necessary.

Furthermore, because of a technical mistake, some participants who completed *VISOR* but did not complete t=1 (directly after the DA) did not receive an automatic invite for the other follow-ups. When this mistake was discovered, some participants were already lost to follow-up. We dealt with this in two different ways: (1) people who had already missed t=2 but completed the DA in the 3 months before the discovery of the mistake received the invite to participate in t=2 regardless (38/599, 6.3% received this invitation and 3/599, 0.5% made use of this), and (2) people who had already missed t=3 were still invited to participate in t=3 but not t=2 (130/599, 21.7% received the invitation and 5/599, 0.8% made use of this).

Finally, we originally planned to adjust our analyses for covariates that were selected a priori if they were associated with outcomes. However, based on the advice of the involved statistician (SJ), we decided to adjust our analyses for all selected covariates a priori (as described in the study protocol [28]) to keep the covariates consistent across the different models and make the models comparable.

Results

Sample

The total sample comprised 2375 participants who were randomized, of whom 1164 (49.01%) completed the baseline questionnaire. Subsequently, of the 2375 participants, 599 (25.22%) completed one of the DAs, 276 (11.62%) filled in t=1 completely, 97 (4.08%) filled in t=2 completely, and 103 (4.34%) filled in t=3 completely. The entire trial flow is shown in Figure 3 [57]. The characteristics of the participants who completed the baseline questionnaire are shown in Table 1.

Figure 3. Trial flow adapted from the CONSORT (Consolidated Standards of Reporting Trials) Flow Diagram [57]—the number of participants that were used for the data analyses can be found in the subsequent tables. t=1: immediately after the use of the DA; t=2: 1 month after the use of the DA; t=3: 6 months after the use of the DA; *30 accounts showed duplicate email addresses; of those, only 8 were linked to accounts in both trial arms (ie, multiple accounts using the same email address in each of the trial arms) and finished baseline with >1 account; of those only 2 participants filled in the follow-ups in such a way that they could have a distorting effect on the results (ie, they were first randomized to the intervention group, then to the control group, and then filled in the follow-ups as participants belonging to the control group, although they received the additional intervention elements); therefore, those were adjusted (ie, the randomization variable was changed to 1=intervention). DA: decision aid; t=1: time point 1; t=2: time point 2; t=3: time point 3.

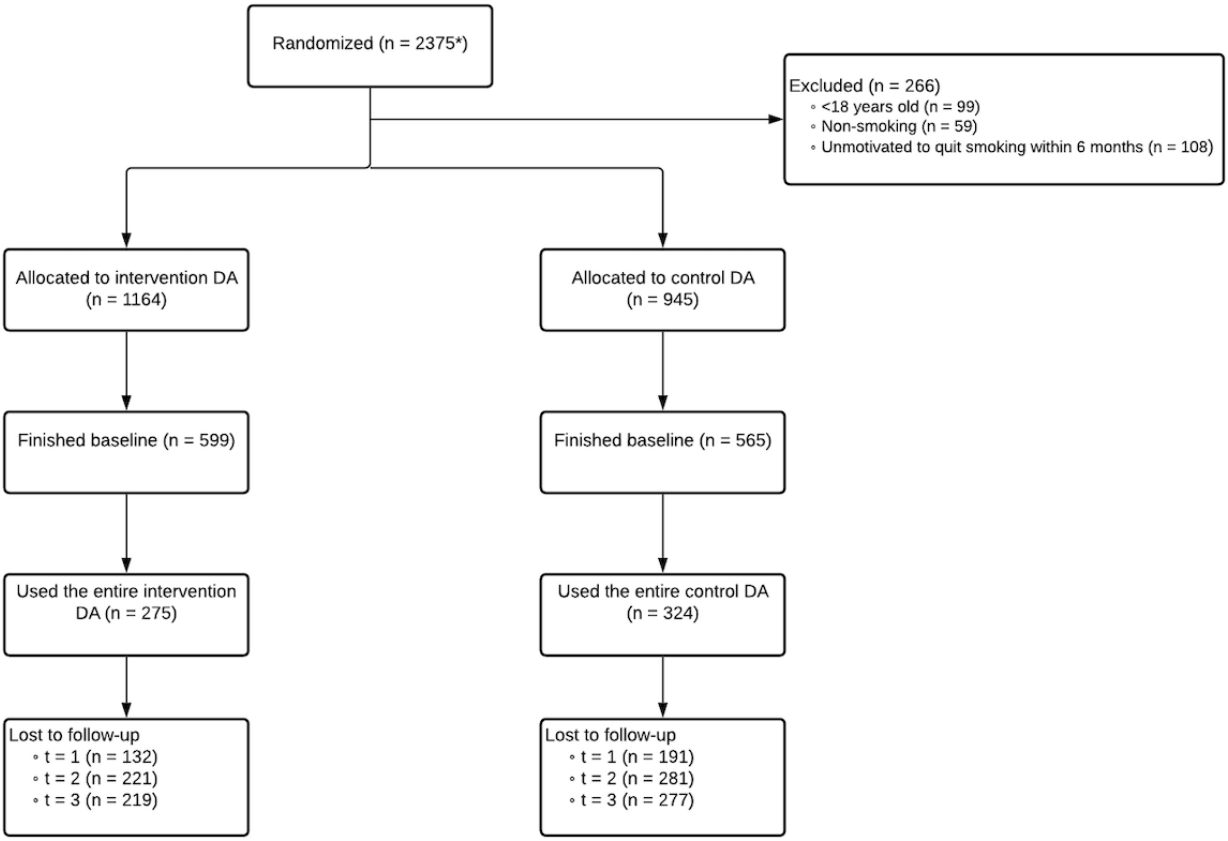


Table 1. Characteristics of participants who finished the baseline questionnaire (N=1164).

Participant characteristics	Entire sample	Intervention group (n=599)	Control group (n=565)
Gender, n (%)			
Women	738 (63.4)	372 (62.1)	366 (64.8)
Men	424 (36.4)	225 (37.6)	199 (35.2)
Nonbinary	1 (0.09)	1 (0.2)	0 (0.0)
Prefers not to say	1 (0.09)	1 (0.2)	0 (0.0)
Age (years), n (%)			
18-23	326 (28.01)	173 (28.9)	153 (27.1)
24-29	155 (13.32)	73 (12.2)	82 (14.5)
30-100	683 (58.68)	353 (58.9)	330 (58.4)
Education, n (%)			
Low	151 (12.97)	88 (14.7)	63 (11.2)
Medium	661 (56.78)	321 (53.6)	340 (60.2)
High	352 (30.24)	190 (31.7)	162 (28.7)
Tobacco products[a], n (%)			
Cigarettes	1144 (98.28)	586 (97.8)	558 (98.8)
E-cigarettes[b]	56 (4.81)	28 (4.7)	28 (5)
Pipe	6 (0.52)	4 (0.7)	2 (0.4)
Cannabis	42 (3.61)	23 (3.8)	19 (3.4)
Cigar	17 (1.46)	10 (1.7)	7 (1.2)
Other	16 (1.37)	8 (1.3)	8 (1.4)
Tobacco consumption			
Total without e-cigarettes (daily), mean (SD)	16.12 (8.64)	16.22 (9.09)	16.00 (8.13)
e-Cigarettes only[c], n (%)			
Less than monthly	6 (10.7)	6 (21.4)	0 (0.0)
Less than weekly but at least once per month	10 (17.9)	2 (7.1)	8 (28.6)
Less than daily but at least once per week	12 (21.4)	7 (25.0)	5 (17.9)
Daily but not multiple times	4 (7.1)	3 (10.7)	1 (3.6)
Multiple times per day	24 (42.9)	10 (35.7)	14 (50.0)
Smoking cessation behavior			
Ever smoking cessation attempt, n (%)	1032 (88.66)	522 (87.1)	510 (90.3)
Smoking cessation attempts (lasting 24 hours), mean (SD)[d]	4.19 (8.76)	3.99 (9.31)	4.40 (8.16)
Cessation assistance use in the past 6 months[e], n (%)			
Evidence based	169 (14.52)	82 (13.7)	87 (15.4)
Nonevidence based	23 (1.98)	6 (1.0)	17 (3.0)
Stage of decision-making			
Has not begun to think about the choices, n (%)	185 (15.89)	91 (15.2)	94 (16.6)
Has not begun to think about the choices but is interested in doing so, n (%)	288 (24.74)	147 (24.5)	141 (25.0)
Is considering the options now, n (%)	404 (34.71)	211 (35.2)	193 (34.2)
Is close to selecting an option, n (%)	87 (7.47)	45 (7.5)	42 (7.4)
Already made a decision but is still willing to reconsider, n (%)	124 (10.65)	69 (11.5)	55 (9.7)

Participant characteristics	Entire sample	Intervention group (n=599)	Control group (n=565)
Has already made a decision and is unlikely to change their mind, n (%)	76 (6.53)	36 (6.0)	40 (7.1)
Values, mean (SD)	2.92 (1.4)	2.94 (1.39)	2.90 (1.42)
FTND-R[f], mean (SD)	6.66 (3.30)	6.59 (3.35)	6.75 (3.25)

[a]Selecting multiple products was possible.

[b]All dual users.

[c]Percentages refer to e-cigarette users only.

[d]Excluding extreme outliers ≥1000 and participants who had never attempted to stop smoking before.

[e]At least one, can be multiple; percentages >100% are because of rounding.

[f]FTND-R: Revised Fagerström Test for Nicotine Dependence.

Attrition

Nonuse Attrition

Comparisons of the participants who did not complete one of the DAs and those who did showed significant differences in group allocation (χ^2_1 [N=1164]=15.2; P<.001), age (χ^2_2 [N=1164]=78.3; P<.001), level of education χ^2_2 [N=1164]=12.5; P=.002), whether people used other tobacco products (χ^2_1 [N=1164]=4.6; P=.03), the average number of cessation attempts (*W*=121,312; P=.02), and stage of decision-making at baseline (*W*=153,400; P=.004). All comparisons can be found in Multimedia Appendix 5. In the logistic regression model, having been allocated to the intervention group (odds ratio [OR] 1.65, 95% CI 1.28-2.14; P<.001), being aged between 24 and 29 years (compared with 18-23 years; OR 0.60, 95% CI 0.38-0.93; P=.02), being aged between 30 and 100 years (compared with 18-23 years; OR 0.29, 95% CI 0.21-0.39; P<.001), and a high (compared with a low) level of education (OR 0.64, 95% CI 0.42-0.99; P=.046) remained significant; that is, participants in the intervention group were more likely *not* to complete *VISOR*, whereas those aged 24 to 100 years (compared with 18-23 years) and those with a high (compared with a low) level of education were more likely to complete *VISOR*.

Dropout Attrition

Comparisons of the participants who did not complete t=1 and those who did showed significant differences in group allocation (χ^2_1 [N=599]=7.2; P=.01), gender (χ^2_1 [N=598]=5.9; P=.02), and stage of decision-making (*W*=37,942; P=.001). All comparisons can be found in Multimedia Appendix 6. All variables remained significant in the logistic regression model: having been allocated to the intervention group (OR 0.65, 95% CI 0.47-0.90; P=.01), men compared with women (OR 0.65, 95% CI 0.46-0.92; P=.01), and stage of decision-making (OR 0.82, 95% CI 0.72-0.92; P=.001); that is, participants in the intervention group, men (compared with women) and those in higher stages of decision-making were less likely to drop out.

Comparisons of the participants who did not complete t=2 and those who did showed significant differences in group allocation (χ^2_1 [N=599]=4.4; P=.04) and stage of decision-making (*W*=21,158; P=.03). All comparisons can be found in Multimedia Appendix 7. In the logistic regression model, having been allocated to the intervention group remained significant (OR 0.63, 95% CI 0.41-0.98; P=.04); that is, participants in the intervention group were less likely to drop out.

Comparisons of the participants who did not complete t=3 and those who did showed significant differences only in the stage of decision-making (*W*=21,601; P=.01), which also remained significant in the logistic regression (OR 0.84, 95% CI 0.72-0.97; P=.02); that is, participants in higher stages of decision-making were less likely to drop out. All comparisons can be found in Multimedia Appendix 8.

As next to the demographic variables (which were already planned as covariates for all analyses) and group allocation, only the stage of decision-making was most consistently associated with dropout; we decided to include the stage of decision-making as a covariate for all analyses as well—and not only decisional conflict as planned in the protocol [28].

Hypotheses Testing

Hypothesis 1a: Smoking Cessation After 1 Month

Although it was observed that more participants stopped smoking in the intervention group (15/55, 27% of the respondents) than in the control group (7/46, 15% of the respondents) after 1 month, the intervention did not result in a significant effect on smoking cessation in the complete case analyses (OR 2.09, 95% CI 0.79 to infinity [+], crude P=.07; OR 1.93, 95% CI 0.64 to infinity [+], adjusted P=.13) or the MI analyses (OR 1.26, 95% CI 0.63 to infinity [+], crude P=.25; OR 1.35, 95% CI 0.57 to infinity [+], adjusted P=.24). However, in the worst-case scenario, effects in favor of the intervention group were observed (OR 2.61, 95% CI 1.08 to infinity [+], crude P=.02; OR 2.71, 95% CI 1.09 to infinity [+], adjusted P=.02). In 2 of the 3 adjusted models (ie, complete cases and worst-case scenario), the only (other) variable that had a significant effect on smoking cessation was the stage of decision-making (complete case analysis: OR 1.83, 95% CI 1.24-2.85; P=.004; worst-case scenario: OR 1.63, 95% CI 1.22-2.22; P=.001). In the MI scenario, none of the included variables had a significant effect on smoking cessation rates after 1 month. Therefore, hypothesis 1a could only be confirmed in the worst-case scenario. Table 2 presents the results and more information.

Table 2. Results of logistic regression for hypothesis 1a and 1b: smoking cessation after 1 month and 6 months.

Time point and variable	B (SE)	P value	Odds ratio (95% CI)
After 1 month			
Complete cases (crude)[a]			
Intercept	−1.72 (0.41)	<.001	N/A[b]
Group allocation (intervention)	0.74 (0.51)	.07[c]	2.09 (0.79 to infinity)
Complete cases (adjusted)[d]			
Intercept	−3.30 (1.67)	.048	N/A
Group allocation (intervention)	0.66 (0.58)	.13[c]	1.93 (0.64 to infinity)
Age (24-29 years)	−1.53 (1.30)	.24	0.22 (0.01 to 2.13)
Age (30-100 years)	−1.27 (0.70)	.07	0.28 (0.07 to 1.12)
Gender (men)	0.21 (0.57)	.72	1.23 (0.39 to 3.76)
Education (medium)	0.01 (1.12)	.99	1.01 (0.12 to 10.67)
Education (high)	0.36 (1.11)	.74	1.44 (0.17 to 14.87)
FTND-R[e]	0.04 (0.08)	.65	1.04 (0.88 to 1.22)
Stage of decision-making	0.61 (0.21)	.004	1.83 (1.24 to 2.85)
Worst-case scenario (crude)[f]			
Intercept	−3.81 (0.38)	<.001	N/A
Group allocation (intervention)	0.96 (0.47)	.02[c]	2.61 (1.08 to infinity)
Worst-case scenario (adjusted)[g]			
Intercept	−5.21 (1.23)	<.001	N/A
Group allocation (intervention)	1.0 (0.48)	.02[c]	2.71 (1.09 to infinity)
Age (24-29 years)	−1.75 (1.11)	.11	0.17 (0.01 to 1.08)
Age (30-100 years)	−0.94 (0.53)	.08	0.39 (0.14 to 1.17)
Gender (men)	0.03 (0.47)	.94	1.04 (0.40 to 2.53)
Education (medium)	0.32 (0.83)	.69	1.38 (0.32 to 9.64)
Education (high)	0.89 (0.84)	.29	2.44 (0.55 to 17.32)
FTND-R	−0.01 (0.07)	.86	0.99 (0.86 to 1.13)
Stage of decision-making	0.49 (0.15)	.001	1.63 (1.22 to 2.22)
Multiple imputations (crude)[h]			
Intercept	−0.82 (0.43)	.07	N/A
Group allocation (intervention)	0.23 (0.35)	.25[c]	1.26 (0.63 to infinity)
Multiple imputations (adjusted)[i]			
Intercept	−1.66 (1.41)	.24	N/A
Group allocation (intervention)	0.30 (0.43)	.24[c]	1.35 (0.57 to infinity)
Age (24-29 years)	−0.95 (1.00)	.35	0.39 (0.05 to 2.88)
Age (30-100 years)	−0.75 (0.59)	.21	0.47 (0.15 to 1.54)
Gender (men)	−0.06 (0.51)	.90	0.94 (0.34 to 2.62)
Education (medium)	0.01 (0.99)	.99	1.01 (0.14 to 7.40)
Education (high)	0.33 (0.96)	.74	1.39 (0.20 to 9.64)
FTND-R	0.02 (0.06)	.72	1.02 (0.90 to 1.16)

Time point and variable	B (SE)	P value	Odds ratio (95% CI)
Stage of decision-making	0.35 (0.21)	.10	1.42 (0.93 to 2.17)
After 6 months			
Complete cases (crude)[j]			
Intercept	−1.0 (0.31)	<.01	N/A
Group allocation (intervention)	0.45 (0.41)	.14[c]	1.56 (0.71 to infinity)
Complete cases (adjusted)[k]			
Intercept	−0.50 (1.09)	.65	N/A
Group allocation (intervention)	0.24 (0.45)	.30[c]	1.27 (0.52 to infinity)
Age (24-29 years)	−1.82 (1.19)	.13	0.16 (0.01 to 1.24)
Age (30-100 years)	−0.49 (0.58)	.40	0.61 (0.19 to 1.94)
Gender (men)	0.74 (0.45)	.10	2.09 (0.87 to 5.06)
Education (medium)	−0.80 (0.75)	.29	0.45 (0.10 to 2.04)
Education (high)	−0.46 (0.78)	.55	0.63 (0.13 to 2.98)
FTND-R	−0.10 (0.07)	.12	0.90 (0.78 to 1.03)
Stage of decision-making	0.31 (0.16)	.05	1.36 (1.00 to 1.88)
Worst-case scenario (crude)[l]			
Intercept	−3.10 (0.27)	<.001	N/A
Group allocation (intervention)	0.70 (0.35)	.02[c]	2.02 (1.03 to infinity)
Worst-case scenario (adjusted)[m]			
Intercept	−3.04 (0.86)	<.001	N/A
Group allocation (intervention)	0.62 (0.36)	.04[c]	1.85 (0.92 to infinity)
Age (24-29 years)	−1.80 (1.08)	.10	0.17 (0.01 to 0.96)
Age (30-100 years)	−0.17 (0.44)	.69	0.84 (0.36 to 2.09)
Gender (men)	0.64 (0.35)	.07	1.90 (0.95 to 3.79)
Education (medium)	−0.36 (0.56)	.52	0.70 (0.25 to 2.32)
Education (high)	−0.13 (0.58)	.82	0.88 (0.29 to 3.0)
FTND-R	−0.14 (0.06)	.02	0.87 (0.78 to 0.98)
Stage of decision-making	0.32 (0.12)	.01	1.37 (1.09 to 1.72)
Multiple imputations (crude)[n]			
Intercept	−0.78 (0.31)	.01	N/A
Group allocation (intervention)	0.30 (0.41)	.23[c]	1.35 (0.59 to infinity)
Multiple imputations (adjusted)[o]			
Intercept	0.14 (1.06)	.89	N/A
Group allocation (intervention)	0.19 (0.43)	.33[c]	1.21 (0.52 to infinity)
Age (24-29 years)	−1.66 (1.07)	.13	0.19 (0.02 to 1.62)
Age (30-100 years)	−0.45 (0.45)	.32	0.64 (0.26 to 1.56)
Gender (men)	0.65 (0.50)	.20	1.91 (0.70 to 5.21)
Education (medium)	−0.98 (0.72)	.18	0.37 (0.09 to 1.57)
Education (high)	−0.56 (0.75)	.46	0.57 (0.13 to 2.57)
FTND-R	−0.10 (0.06)	.09	0.91 (0.81 to 1.02)

Time point and variable	B (SE)	P value	Odds ratio (95% CI)
Stage of decision-making	0.22 (0.16)	.18	1.24 (0.90 to 1.71)

[a]R^2=0.02 (Hosmer-Lemeshow), 0.02 (Cox-Snell), 0.03 (Nagelkerke); χ^2_1=2.2; P=.14; 101/1164, 8.68%.

[b]N/A: not applicable.

[c]1-sided.

[d]R^2=0.14 (Hosmer-Lemeshow), 0.13 (Cox-Snell), 0.21 (Nagelkerke); χ^2_7=12.3; P=.09 (compared with the crude model), 101/1164, 8.68%

[e]FTND-R: Revised Fagerström Test for Nicotine Dependence.

[f]R^2=0.02 (Hosmer-Lemeshow), 0.01 (Cox-Snell), 0.03 (Nagelkerke); χ^2_1=4.6; P=.03; 599/1164, 51.46%.

[g]R^2=0.11 (Hosmer-Lemeshow), 0.03 (Cox-Snell), 0.13 (Nagelkerke); χ^2_7=16.2; P=.02 (compared with the crude model, excluding the nonbinary participant), 598/1164, 51.37%.

[h]R^2=0.01 (Hosmer-Lemeshow), 0.01 (Cox-Snell), 0.01 (Nagelkerke); χ^2_1=4.4; P=.52; 599/1164, 51.46%.

[i]R^2=0.1 (Hosmer-Lemeshow), 0.11 (Cox-Snell), 0.16 (Nagelkerke); χ^2_7=70.1; P=.58 (compared with the crude model, excluding the nonbinary participant), 598/1164, 51.37%.

[j]R^2=0.01 (Hosmer-Lemeshow), 0.01 (Cox-Snell), 0.01 (Nagelkerke); χ^2_1=1.21; P=.27; 115/1164, 9.88%.

[k]R^2=0.11 (Hosmer-Lemeshow), 0.12 (Cox-Snell), 0.17 (Nagelkerke); χ^2_7=14.1; P=.05 (compared with the crude model), 115/1164, 9.88%.

[l]R^2=0.02 (Hosmer-Lemeshow), 0.01 (Cox-Snell), 0.02 (Nagelkerke); χ^2_1=4.2; P=.04; 599/1164, 51.46%.

[m]R^2=0.09 (Hosmer-Lemeshow), 0.04 (Cox-Snell), 0.11 (Nagelkerke); χ^2_7=20.9; P<.01 (compared with the crude model, excluding the nonbinary participant), 598/1164, 51.37%.

[n]R^2=0.01 (Hosmer-Lemeshow), 0.01 (Cox-Snell), 0.02 (Nagelkerke); χ^2_1=7.2; P=.48; 599/1164, 51.46%.

[o]R^2=0.12 (Hosmer-Lemeshow), 0.14 (Cox-Snell), 0.19 (Nagelkerke); χ^2_7=77.3;P=.24 (compared with the crude model, excluding the nonbinary participant), 598/1164, 51.37%.

Hypothesis 1b: Smoking Cessation After 6 Months

Although it was again observed that after 6 months, more participants reported having stopped smoking in the intervention group (23/63, 37% of the respondents) than in the control group (14/52, 27% of the respondents), the intervention did not result in a significant effect on smoking cessation in the complete case analyses (OR 1.56, 95% CI 0.71 to infinity [+], crude P=.14; OR 1.27, 95% CI 0.52 to infinity [+], adjusted P=.30) or the MI analyses (OR 1.35, 95% CI 0.59 to infinity [+], crude P=.23; OR 1.21, 95% CI 0.52 to infinity [+], adjusted P=.33). Similar to the results based on data collected after 1 month, effects in favor of the intervention group were observed in the worst-case scenario (OR 2.02, 95% CI 1.03 to infinity [+], crude P=.02; OR 1.85, 95% CI 0.92 to infinity [+], adjusted P=.04). In the worst-case scenario, the FTND-R (OR 0.87, 95% CI 0.78-0.98; P=.02; ie, nicotine dependence) and stage of decision-making (OR 1.37, 95% CI 1.09-1.72; P=.01) had a significant effect on smoking cessation rates. However, the FTND-R violated the linearity assumption in this model, but excluding it did not change the conclusions in relation to the primary outcome (Multimedia Appendix 9). In the complete case analyses and the MI scenario, none of the included variables had a significant effect on smoking cessation rates after 6 months. Therefore, hypothesis 1b could only be confirmed in the worst-case scenario. Table 2 presents the results and more information.

Hypothesis 2a: Use of Evidence-Based Cessation Assistance After 1 Month

Although, after 1 month, more people in the intervention group reported to have used an evidence-based cessation assistance tool (26/56, 46% of the respondents) than those in the control group (18/47, 38% of the respondents), the intervention did not result in a significant effect on the uptake of evidence-based cessation assistance in the complete case analysis (OR 1.40, 95% CI 0.64 to infinity [+], crude P=.20; OR 1.42, 95% CI 0.60 to infinity [+], adjusted P=.21) or the MI analyses (OR 1.25, 95% CI 0.69 to infinity [+], crude P=.23; OR 1.42, 95% CI 0.61 to infinity [+], adjusted P=.20). Similar to smoking cessation, effects in favor of the intervention group appeared in the worst-case scenario but only in the crude model (OR 1.78, 95% CI 0.96 to infinity [+], crude P=.04; OR 1.68, 95% CI 0.89 to infinity [+], adjusted P=.05). In the adjusted models, only the stage of decision-making had a significant effect on the outcome (in the complete case analysis, OR 1.53, 95% CI 1.11-2.19, P=.01; in the worst-case scenario, OR 1.36, 95% CI 1.10-1.68; P=.005; in the MI analysis, OR 1.47, 95% CI 1.07-2.02; P=.02). Therefore, hypothesis 2a could only be partially confirmed in the worst-case scenario. Table 3 presents the results and more information.

Table 3. Results of logistic regression for hypothesis 2a and 2b: use of evidence-based cessation assistance after 1 month and 6 months.

Time point and variable	B (SE)	P value	Odds ratio (95% CI)
After 1 month			
Complete cases (crude)[a]			
Intercept	−0.48 (0.30)	.11	N/A[b]
Group allocation (intervention)	0.33 (0.40)	.20[c]	1.40 (0.64 to infinity)
Complete cases (adjusted)[d]			
Intercept	−2.76 (0.85)	<.01	N/A
Group allocation (intervention)	0.35 (0.44)	.21[c]	1.42 (0.60 to infinity)
Age (24-29 years)	0.49 (0.96)	.61	1.62 (0.23 to 10.65)
Age (30-100 years)	0.51 (0.61)	.40	1.66 (0.52 to 5.74)
Gender (men)	0.10 (0.46)	.82	1.11 (0.45 to 2.76)
Education (high)	−0.29 (0.45)	.52	0.75 (0.30 to 1.80)
FTND-R[e]	0.08 (0.06)	.19	1.09 (0.96 to 1.24)
Stage of decision-making	0.43 (0.17)	.01	1.53 (1.11 to 2.19)
Worst-case scenario (crude)[f]			
Intercept	−2.83 (0.24)	<.001	N/A
Group allocation (intervention)	0.57 (0.32)	.04[c]	1.78 (0.96 to infinity)
Worst-case scenario (adjusted)[g]			
Intercept	−3.98 (0.82)	<.001	N/A
Group allocation (intervention)	0.52 (0.32)	.05[c]	1.68 (0.89 to infinity)
Age (24-29 years)	−0.34 (0.74)	.65	0.71 (0.14 to 2.91)
Age (30-100 years)	0.24 (0.48)	.62	1.27 (0.53 to 3.56)
Gender (men)	0.05 (0.33)	.89	1.05 (0.54 to 1.98)
Education (medium)	−0.33 (0.48)	.49	0.72 (0.29 to 1.97)
Education (high)	−0.06 (0.50)	.90	0.94 (0.36 to 2.66)
FTND-R	0.03 (0.05)	.57	1.03 (0.93 to 1.14)
Stage of decision-making	0.30 (0.11)	.005	1.36 (1.10 to 1.68)
Multiple imputations (crude)[h]			
Intercept	−0.55 (0.34)	.11	N/A
Group allocation (intervention)	0.22 (0.30)	.23[c]	1.25 (0.69 to infinity)
Multiple imputations (adjusted)[i]			
Intercept	−2.71 (0.86)	<.01	N/A
Group allocation (intervention)	0.35 (0.42)	.20[c]	1.42 (0.61 to infinity)
Age (24-29 years)	0.33 (0.86)	.70	1.39 (0.25 to 7.72)
Age (30-100 years)	0.45 (0.52)	.39	1.57 (0.56 to 4.43)
Gender (men)	0.12 (0.46)	.79	1.13 (0.45 to 2.85)
Education (high)	−0.20 (0.44)	.65	0.82 (0.34 to 1.97)
FTND-R	0.09 (0.06)	.11	1.10 (0.98 to 1.23)
Stage of decision-making	0.39 (0.16)	.02	1.47 (1.07 to 2.02)
After 6 months			

Time point and variable	B (SE)	P value	Odds ratio (95% CI)
Complete cases (crude)[j]			
Intercept	0.08 (0.28)	.78	N/A
Group allocation (intervention)	0.42 (0.39)	.14[c]	1.53 (0.72 to infinity)
Complete cases (adjusted)[k]			
Intercept	−1.14 (1.13)	.31	N/A
Group allocation (intervention)	0.54 (0.43)	.11[c]	1.71 (0.74 to infinity)
Age (24-29 years)	−1.06 (0.87)	.22	0.35 (0.06 to 1.82)
Age (30-100 years)	0.06 (0.56)	.92	1.06 (0.35 to 3.16)
Gender (men)	−0.15 (0.43)	.73	0.86 (0.37 to 2.01)
Education (medium)	−0.01 (0.78)	.99	0.99 (0.19 to 4.36)
Education (high)	−0.06 (0.81)	.94	0.94 (0.17 to 4.46)
FTND-R	0.10 (0.06)	.12	1.10 (0.98 to 1.25)
Stage of decision-making	0.21 (0.15)	.17	1.24 (0.92 to 1.68)
Worst-case scenario (crude)[l]			
Intercept	−2.44 (0.20)	<.001	N/A
Group allocation (intervention)	0.61 (0.27)	.01[c]	1.84 (1.09 to infinity)
Worst-case scenario (adjusted)[m]			
Intercept	−3.48 (0.70)	<.001	N/A
Group allocation (intervention)	0.59 (0.27)	.02[c]	1.80 (1.06 to infinity)
Age (24-29 years)	−0.94 (0.68)	.17	0.39 (0.08 to 1.34)
Age (30-100 years)	0.10 (0.38)	.80	1.10 (0.54 to 2.40)
Gender (men)	0.02 (0.28)	.94	1.02 (0.58 to 1.76)
Education (medium)	0.27 (0.45)	.55	1.31 (0.57 to 3.44)
Education (high)	0.22 (0.48)	.65	1.25 (0.50 to 3.43)
FTND-R	0.01 (0.04)	.87	1.01 (0.92 to 1.10)
Stage of decision-making	0.24 (0.09)	.01	1.27 (1.06 to 1.53)
Multiple imputations (crude)[n]			
Intercept	−0.09 (0.25)	.73	N/A
Group allocation (intervention)	0.54 (0.37)	.07[c]	1.72 (0.83 to infinity)
Multiple imputations (adjusted)[o]			
Intercept	−1.11 (1.08)	.31	N/A
Group allocation (intervention)	0.50 (0.43)	.13[c]	1.65 (0.69 to infinity)
Age (24-29 years)	−0.84 (0.78)	.29	0.43 (0.09 to 2.06)
Age (30-100 years)	0.15 (0.47)	.76	1.16 (0.45 to 2.96)
Gender (men)	−0.20 (0.47)	.67	0.82 (0.32 to 2.10)
Education (medium)	−0.14 (0.63)	.83	0.87 (0.25 to 3.04)
Education (high)	−0.24 (0.68)	.73	0.79 (0.20 to 3.05)
FTND-R	0.08 (0.06)	.19	1.09 (0.96 to 1.23)

Time point and variable	B (SE)	P value	Odds ratio (95% CI)
Stage of decision-making	0.24 (0.15)	.12	1.27 (0.94 to 1.72)

[a]R^2=0.005 (Hosmer-Lemeshow), 0.01 (Cox-Snell), 0.01 (Nagelkerke); χ^2_1=0.7; P=.41; 103/1164, 8.85%.

[b]N/A: not applicable.

[c]1-sided.

[d]R^2=0.09 (Hosmer-Lemeshow), 0.11 (Cox-Snell), 0.15 (Nagelkerke); χ^2_6=11.7; P=.07 (compared with the crude model), 103/1164, 8.85%.

[e]FTND-R: Revised Fagerström Test for Nicotine Dependence.

[f]R^2=0.01 (Hosmer-Lemeshow), 0.01 (Cox-Snell), 0.01 (Nagelkerke); χ^2_1=3.3; P=.07; 599/1164, 51.46%.

[g]R^2=0.05 (Hosmer-Lemeshow), 0.02 (Cox-Snell), 0.06 (Nagelkerke); χ^2_7=11.2; P=.13 (compared with the crude model, excluding the nonbinary participant), 598/1164, 51.37%.

[h]R^2=0.005 (Hosmer-Lemeshow), 0.01 (Cox-Snell), 0.01 (Nagelkerke); χ^2_1=3.8; P=.46; 599/1164, 51.46%.

[i]R^2=0.11 (Hosmer-Lemeshow), 0.13 (Cox-Snell), 0.18 (Nagelkerke); χ^2_6=77.2; P=.13 (compared with the crude model, excluding the nonbinary participant), 598/1164, 51.37%.

[j]R^2=0.01 (Hosmer-Lemeshow), 0.01 (Cox-Snell), 0.01 (Nagelkerke); χ^2_1=1.2; P=.27; 111/1164, 9.54%.

[k]R^2=0.07 (Hosmer-Lemeshow), 0.09 (Cox-Snell), 0.12 (Nagelkerke); χ^2_7=8.8; P=.26 (compared with the crude model), 111/1164, 9.54%.

[l]R^2=0.01 (Hosmer-Lemeshow), 0.01 (Cox-Snell), 0.02 (Nagelkerke); χ^2_1=5.2; P=.02; 599/1164, 51.46%.

[m]R^2=0.04 (Hosmer-Lemeshow), 0.03 (Cox-Snell), 0.05 (Nagelkerke); χ^2_7=10.8; P=.15 (compared with the crude model, excluding the nonbinary participant), 598/1164, 51.37%.

[n]R^2=0.02 (Hosmer-Lemeshow), 0.02 (Cox-Snell), 0.03 (Nagelkerke); χ^2_1=14.1; P=.14; 599/1164, 51.46%.

[o]R^2=0.10 (Hosmer-Lemeshow), 0.12 (Cox-Snell), 0.16 (Nagelkerke); χ^2_7=65.3; P=.37 (compared with the crude model, excluding the nonbinary participant), 598/1164, 51.37%.

Hypothesis 2b: Use of Evidence-Based Cessation Assistance After 6 Months

Similarly, although, after 6 months, more people in the intervention group reported using an evidence-based cessation assistance tool (38/61, 62% of the respondents) than those in the control group (26/50, 52% of the respondents), the intervention did not result in a significant effect on the uptake of evidence-based cessation assistance in the complete case analysis (OR 1.53, 95% CI 0.72 to infinity [+], crude P=.14; OR 1.71, 95% CI 0.74 to infinity [+], adjusted P=.11) or the MI analysis (OR 1.72, 95% CI 0.83 to infinity [+], crude P=.07; OR 1.65, 95% CI 0.69 to infinity [+], adjusted P=.13). However, in the worst-case scenario, effects in favor of the intervention group emerged (OR 1.84, 95% CI 1.09 to infinity [+], crude P=.01; OR 1.80, 95% CI 1.06 to infinity [+], adjusted P=.02). In the adjusted models, the stage of decision-making had a significant effect on the outcome but only in the worst-case scenario (OR 1.27, 95% CI 1.06-1.53; P=.01). Therefore, hypothesis 2b could only be confirmed in the worst-case scenario. Table 3 presents the results and more information.

Hypothesis 3: Decisional Conflict Immediately After the DA

Despite the small difference in averages between the intervention (mean 39.29, SD 25.00) and control groups (mean 41.17, SD 25.23), the intervention had no significant effect on the decisional conflict in either the complete case analysis (β=−0.04, crude P=.25; β=−0.05, adjusted P=.20) or MI analysis (β=−0.05, crude P=.18; β=−0.05, adjusted P=.22). In the adjusted models, only the stage of decision-making (β=−0.15, P=.005 in the complete case analysis; β=−0.14, P=.01 in the MI analysis), being aged between 30 and 100 years (compared with 18-23 years) (only in the MI analysis, β=−0.06, P=.045), and a high level of education (β=−0.27, P=.002 in the complete case analysis; β=−0.25, P=.005 in the MI analysis) had a significant effect on the outcome. All of them had a negative effect; that is, people who reported a higher stage of decision-making, participants aged between 30 and 100 years old (compared with 18-23 years), and participants with a high level of education (compared with participants with a low level of education) experienced less decisional conflict. Therefore, hypothesis 3 could not be confirmed. Table 4 presents the results and more information.

Table 4. Results of linear regression for hypothesis 3: decisional conflict immediately after using the decision aid[a].

Case analysis	B (95% CI)	SE	β	P value
Complete cases (crude)[b]				
Intercept	41.17 (37.43 to 44.92)	1.90	N/A[c]	<.001
Group allocation (intervention)	−1.89 (infinity to 3.52)	2.75	−0.04	.25[d]
Complete cases (adjusted)[e]				
Intercept	62.69 (50.44 to 74.94)	6.23	N/A	<.001
Group allocation (intervention)	−2.31 (infinity to 3.03)	2.71	−0.05	.20[d]
Age (24-29 years)	−4.05 (−15.08 to 6.98)	5.61	−0.05	.47
Age (30-100 years)	−6.60 (−14.07 to 0.86)	3.80	−0.11	.08
Gender (men)	2.51 (−2.97 to 7.99)	2.79	.05	.37
Education (medium)	−6.44 (−15.12 to 2.24)	4.41	−0.13	.15
Education (high)	−13.91 (−22.85 to −4.98)	4.54	−0.27	.002
Stage of decision-making	−2.69 (−4.55 to −0.84)	0.94	−0.15	.005
Multiple imputations (crude)[f]				
Intercept	42.36 (38.49 to 46.23)	1.96	N/A	<.001
Group allocation (intervention)	−2.53 (infinity to 2.83)	2.71	−0.05	.18[d]
Multiple imputations (adjusted)[g]				
Intercept	62.88 (50.50 to 75.25)	6.25	N/A	<.001
Group allocation (intervention)	−2.32 (infinity to 3.67)	3.02	−0.05	.22[d]
Age (24-29 years)	−4.73 (−14.93 to 5.48)	5.16	−0.06	.36
Age (30-100 years)	−7.47 (−14.78 to −0.16)	3.69	−0.14	.045
Gender (men)	2.27 (−3.26 to 7.79)	2.80	.04	.42
Education (medium)	−6.13 (−15.34 to 3.08)	4.65	−0.12	.19
Education (high)	−12.99 (−21.90 to −4.08)	4.51	−0.25	.005
Stage of decision-making	−2.52 (−4.51 to −0.52)	1.01	−0.14	.01

[a]It should be noted that the residuals were not perfectly normally distributed in the models; this was especially apparent in the crude model (complete cases). However, overall, the skew was not highly substantial.

[b]Multiple R^2=0.001; P=.49; 335/1164, 28.78%.

[c]N/A: not applicable.

[d]1-sided.

[e]Multiple R^2=0.07; P<.001 (compared with the crude model), 335/1164, 28.78%.

[f]Multiple R^2=0.003; P=.36; 599/1164, 51.46%.

[g]Multiple R^2=0.08; P=.001 (compared with the crude model, excluding the nonbinary participant), 598/1164, 51.37%.

Discussion

Principal Findings

The aim of this paper was to report the effects of adding an explicit VCM with computer-tailored advice to a smoking cessation DA (*VISOR*) on both smoking cessation outcomes and decisional conflict. Contrary to our expectations, we did not find any effect on decisional conflict. In addition, although the worst-case scenarios might suggest an effect on smoking cessation rates and cessation assistance uptake, this finding was not replicated in either the complete case analyses or MI analyses. Moreover, given the fact that Blankers et al [58] showed that analyses based on penalized imputation can be biased when missingness is unbalanced between trial arms (as in our case), we cannot confidently speak of the effects on smoking cessation success and cessation assistance uptake, despite the suggestion of effects in the worst-case scenarios. That said, all the significant and nonsignificant effects that were found were in the expected direction; that is, participants in the intervention group showed more smoking cessation, more evidence-based cessation assistance uptake, and less decisional conflict. A summary of the main findings with respect to the primary and secondary outcomes can be found in Textbox 2.

However, based on these findings, it may be too early to declare the addition of explicit VCMs and computer-tailored advice as ineffective to smoking cessation DAs. Previous studies have consistently found effects in favor of explicit VCMs on value-congruent decision-making in other contexts, especially when combined with computer-tailored advice [22,25]. That said, we were the first (to the best of our knowledge) to test this in a smoking cessation context. We would also like to emphasize that our findings do not imply that smoking cessation DAs are ineffective, as we did not compare a smoking cessation DA with usual care, another intervention, or no intervention but rather compared 2 different versions of one and the same DA: one with a VCM and paired computer-tailored advice and one without. In fact, even the control group in this RCT exceeded the smoking cessation rates achieved by the only other Dutch smoking cessation DA described in the literature [17]. However, this study showed that adding explicit VCMs paired with computer-tailored advice might not always be a good idea, as nonuse attrition was 34.19% higher in the intervention group than in the control group. In addition, our nonsignificant findings entail some caveats. In this discussion, we will focus on the two most likely explanations for the lack of effects we found: (1) we were unable to find an effect because of a lack of power, and (2) there truly was no effect. In addition, we also describe the implications of these 2 explanations and how readers might benefit from the insights generated during our project.

Textbox 2. Summary of the main effects.

Summary of the main effects

- Only the scenario in which we assumed that all participants who dropped out continued to smoke suggests that adding an explicit value clarification method and computer-tailored advice to a digital smoking cessation decision aid has statistically significant effects on smoking cessation rates and evidence-based cessation assistance uptake.
- All effects went in the hypothesized directions in all scenarios; that is, in the group that received the additional elements, more participants quit, more participants used evidence-based smoking cessation assistance, and participants experienced less decisional conflict.

Lack of Statistical Power

The most likely explanation for the lack of significant effects was the lack of statistical power. During this trial, we faced a relatively high attrition rate (as is typical for digital interventions [53]), and ultimately, our trial was widely underpowered as even the MI analyses were underpowered to conduct our analyses. On the basis of our power analyses, we planned to include 1592 smokers at baseline, accounting for 50% of attrition. Ultimately, we randomized 2375 participants but had already lost 1211 (50.99%) individuals from this group between randomization and the end of the baseline questionnaire. Within this group, we subsequently lost 48.54% (565/1164) of participants to nonuse attrition and, among the participants invited to follow-up, 53.92% to 83.8% of participants to drop out attrition. Therefore, attrition was much higher than originally anticipated. Although replication of our study with a larger sample is recommended, researchers and DA developers can learn from the reasons for attrition uncovered in this project (eg, use time), as will be explained in the following sections.

Of the allocated participants, only 23.63% (275/1164) of the participants allocated to the intervention group completed *VISOR* compared with 34.29% (324/945) of the participants allocated to the control group. Interestingly, the available data provided by the intervention host indicated that the differences in use times between the 2 groups were significantly different. The median use time of the intervention group was approximately 9 minutes, whereas that of the control group was approximately 6 minutes (see the OSF for more information [35]). This difference may have driven the differences in nonuse attrition between the intervention and control groups. Our other findings regarding nonuse attrition also give credence to this explanation as younger participants were more likely to not complete *VISOR*. During the needs assessment conducted to develop *VISOR*, younger participants in particular indicated that relatively short time frames were acceptable for using a smoking cessation DA [29]. Therefore, future research should explore whether VCMs and computer-tailored advice can be designed to be delivered in a shorter manner to alleviate this problem. For example, Witteman et al [25] reported a digital VCM (including computer-tailored advice) that comprised dynamic web sliders representing both values and preferences (ie, the included options). These web sliders were then linked, meaning that for participants who indicated that a particular value was very important to them, the slider representing the preference moved equally. Such a VCM could potentially be much shorter to use as end users do not have to answer multiple statements. However, at this point, this technique is relatively difficult to use for decisions involving multiple options, which is why we were unable to use it within *VISOR*. However, on the basis of our findings, one might also conclude that it is important to spend more time studying user experience components, especially use times. A well-known method of studying this is the think-aloud method (which has also been used to test *VISOR*'s usability [28]). In studies using the think-aloud method, participants are asked to use an intervention while verbalizing their thoughts to uncover the cognitive processes and emotional reactions when using the intervention [59,60].

Interestingly, based on our data, it can also be concluded that we did not experience a recruitment problem (ie, 2109 participants were initially randomized and eligible to participate) but rather a problem of retention—only a small number of participants completed *VISOR* and even fewer participants completed the follow-up questionnaires. Although this is often observed in digital health care interventions [61], it is crucial to find ways of increasing the actual use of digital DAs and retention in DA trials to ensure benefits for users and also reach sufficient statistical power during studies. A way of achieving this would be to embed DAs (such as *VISOR*) in a counseling pathway as the involvement of a professional has been shown to positively influence time spent on websites aimed at

improving healthy lifestyles [61]. Concurrently, this might also positively influence retention rates in studies (ie, people returning for follow-up measurements) in which digital DAs are evaluated [62,63]. There could be multiple ways of accomplishing this in the context of *VISOR* (or other digital DAs for that matter): (1) *VISOR* could be used together with a health care professional (eg, a practice nurse) or could be sent to participants before or after a health care consultation in which (the result of) *VISOR* is discussed, or (2) a digital form of counseling could be included in *VISOR* (eg, through the form of videocalls [64] or an automated chatbot [65]). The second approach might be especially promising as it would keep the fully digital and automated nature of *VISOR* intact, thereby still ensuring the optimal reach of the DA. Future research should investigate which modalities are especially beneficial in terms of outcomes and retention and which modalities are preferred by end users themselves.

Related to the problem of retention, an interesting chance finding of our study was that it highlighted the importance of the stage of decision-making in this regard, as this influenced both nonuse and dropout attrition. Interestingly, this influence seemed to disappear (partially) once we corrected for other dropout predictors. The most plausible explanation for this is that different groups within our sample (eg, older compared with younger participants) differed in their stage of decision-making (see the OSF for more information [35]), leading to the stage of decision-making becoming insignificant once all predictors were added in the same model. Interestingly, the stage of decision-making is much less routinely assessed than decisional conflict; to illustrate, the validation article of the original Decisional Conflict Scale [9] has been cited >2000 times according to Google Scholar, whereas the user manual of the stage of decision-making scale [50] has only been cited a little over 30 times and, when it is assessed, it is mostly done to include it as a covariate [66,67]. Consequently, the roles of these decisional stages seem to be much less understood. That said, it is assumed that individuals' stages of decision-making influence their receptiveness to DAs and that people who are in active stages of deliberation would benefit the most from DAs, whereas people who either have not even begun to think about the decision or are unwilling to reconsider their decision may not benefit as much from a DA [50]. Interestingly, the largest group (approximately 40%) in our sample was at an early stage of decision-making, which is also reflected in the respondents' average score for this variable (ie, 2.92, SD 1.4), which is slightly below active deliberation (ie, the stage in which individuals start weighing the different options). As such, it might be the case that *VISOR* (with or without the explicit VCM and advice) simply overwhelmed a big part of our sample. The relatively high decisional conflict scores in both groups seem to confirm this. Therefore, tailoring DAs to the stage of decision-making might be a promising approach to further limiting attrition. Although demographic variables seem to be the more obvious choice based on our data, the 1-item scale used to assess the stage of decision-making has the major advantage of being easy to use and not requiring participants to complete lengthy questionnaires. By tailoring DAs to the stage of decision-making (as opposed to other constructs or demographic characteristics), it would be possible to alleviate one of the core problems of most contemporary approaches to computer tailoring—that they often impose a significant burden on participants [68]. The fact that participants' stage of decision-making had the most consistent effect on the outcomes among the included predictors only adds to this. For example, individuals in the initial stages could be offered interventions that are merely educational and supportive in nature, whereas individuals in active deliberation stages could be offered traditional DAs, and individuals in later stages could be offered support to implement their decision. To illustrate, individuals in an early decision-making stage could be offered a brief intervention aimed at increasing motivation to stop smoking or use evidence-based cessation assistance (similar to the 5 As [69]). Once they have reached a higher stage, they could then be provided with a traditional DA, such as *VISOR*. In other words, insights from behavior change could be used to further improve DAs for specific groups of decision makers [70].

True Absence of an Effect

Owing to the lack of statistical power, we cannot confidently conclude that there truly was no effect of *VISOR*. However, interestingly, Sheridan et al [71] also tested the additional effect of an explicit VCM added to a DA aimed at a prevention-related decision (ie, heart disease prevention) and found no effect on outcomes such as decisional conflict or intention to reduce heart disease risk. However, most other studies tested DAs (and by extension explicit VCMs) in a treatment context (eg, between different surgeries [25]), with positive effects reported quite often. On the basis of this contradiction, we could deduce that prevention-related decisions are somewhat different and may not be affected by DAs in a manner similar to treatment decisions. Anecdotally, throughout the project in which *VISOR* was developed and tested, people not involved in the project often stated that prevention-related decisions (such as smoking cessation decisions) seemed easier to make than other decisions, such as treatment decisions. Although this conclusion is logical to a certain degree, as treatment decisions often involve much more imminent risks, such as death in the near future [25], we found no indications that the participants in this study regarded the decision among different cessation aids as easy. In fact, participants in both trial arms experienced decisional conflict that scored above the accepted cutoff value of 37.5 [51]. The decisional conflict scores observed in this study were also higher than those reported in the literature [72], which tend to focus on treatment decisions, indicating that prevention-related decisions are not necessarily easier to make than treatment decisions. In other words, there is no indication that prevention-related decisions are perceived differently from treatment decisions (at least not in terms of experienced difficulty), meaning that they can be assumed to respond similarly to intervention elements commonly included in DAs (such as VCMs). However, based on our data, it cannot be fully excluded that prevention-related decisions differ in factors other than experienced difficulty and therefore do not respond to explicit VCMs as expected. A more viable difference between prevention-related DAs and DAs focused on treatment decisions is that those focused on prevention are often used without the direct involvement of a health care professional [16,17]; that is, prevention-related DAs are more likely to be

self-administered [73] than DAs focused on treatments. However, a recent systematic review and meta-analysis by Larsen et al [73] showed that self-administered DAs for colorectal cancer screening can also be beneficial in relation to prevention-related decisions, such as population screening, showcasing that this explanation (ie, lack of health care involvement) is unlikely. In other words, at this point, no convincing argument can be made about prevention-related decisions being different from other health care decisions. Therefore, they should respond similarly to the elements typically used in DAs. However, given the substantial amount of heterogeneity in the literature on DAs aimed at primary prevention [74], it may be worth investigating further whether prevention-related decisions really respond to DA elements, as one would expect based on the wider DA literature; for example, by replicating our work with a larger sample. In addition, more research could be conducted on perceived differences between prevention-related decisions and other health-related decisions (eg, treatment decisions); for example, by conducting in-depth interviews with participants who recently took both to discover perceived differences in their decision-making.

Strengths and Limitations

The major strength of our study was that it was conducted in a *real-life* context, and we mainly recruited individuals who were interested in smoking cessation (as opposed to the hypothetical scenario often used to evaluate VCMs [25]). In addition, our sample was relatively representative of the Dutch population, except for women being overrepresented and participants with a low level of education being slightly underrepresented [75]. Thus, our findings might be largely generalizable to the smoking Dutch population with access to the internet. The fact that both groups received a DA also allowed us to blind participants, which strengthened the RCT, as participants were not biased in the sense that they knew to which intervention arm they were allocated and, thus, what the other group was offered. Finally, we focused our analyses on 3 outcomes of interest, thereby decreasing the risk of a type I error.

However, this study also has some limitations. Owing to our study design, we might have experienced a selection bias as we only invited people to complete the follow-ups if they had used the entire intervention. We tried to alleviate the influence of this selection bias by including variables associated with this selection (eg, age) as covariates in the analysis and imputation models. However, it is possible that more motivated smokers completed the intervention and therefore, for example, showed higher quit rates. In addition, we had to change the primary outcome and shorten the follow-up period to 6 months after baseline. Although unfortunate, we do not consider this to be a major issue as experts regard a follow-up period of 6 months to be sufficient, as most relapses occur relatively soon after the quit attempt [76-78]. In addition, the trial showed such large attrition that we were underpowered for most of our analyses. That said, the findings from the complete case analysis and those based on MI led to the same conclusions, making our findings more robust. It might also be a limitation that we relied on self-reported data. However, as previous research has shown that self-reported smoking status tends to be accurate [79] and because participants received the reward for study completion regardless of their smoking status, we are confident that this has had little to no influence on our findings. Finally, we were unable to separately test the effects of explicit VCM and computer-tailored advice. This would have required 3 trial arms and an even larger sample size and was therefore deemed unfeasible. It should also be recognized that digital, web-based DAs, by definition, exclude people who either do not have access to the internet at all or do not have the necessary digital literacy to use it. Although we consider this limitation to be small in the Dutch context, given the large number of Dutch households with internet access [80], it is certainly important to consider this to achieve full health equity. In the same vein, it is also important to consider other characteristics (eg, health literacy and financial resources to access smoking cessation assistance) that may influence the ability of end users to benefit from smoking cessation DAs. For example, the findings of our attrition analyses show that future projects should place a stronger emphasis on end users with less formal education. Finally, this study focused on direct behavioral and decisional effects only, as indirect effects were beyond the scope of this study. The same applies to cost-effectiveness and cost-utility. Investigating indirect effects (as hypothesized in the study protocol [28]), cost-effectiveness, and cost-utility might provide additional insights, in addition to the direct effects reported in this study.

Conclusions

At this point, we cannot confidently recommend the inclusion of explicit VCMs and computer-tailored advice in smoking cessation DAs. In fact, these 2 elements might result in higher attrition rates during the use of DAs, thereby limiting their potential. However, our findings in relation to the primary and secondary outcomes might be influenced by a lack of statistical power; therefore, we advocate for the replication of our trial with a larger sample. Finally, researchers should emphasize user experience, especially use times, and experiment with (innovative) solutions to deal with a (too) high perceived user burden, such as the integration of *VISOR* within a counseling pathway or tailoring *VISOR* and other DAs to people's decision-making stages. Finally, researchers should investigate how prevention-related decisions respond to DA elements. Given the strong initial interest in *VISOR* and the potential of smoking cessation DAs to combat tobacco-related mortality by inducing more smoking cessation attempts [15], future research should continue to find ways of optimizing smoking cessation DAs for future use.

Acknowledgments

The authors would like to thank all the participants and individuals who supported us during our recruitment efforts. The developers of the intervention (ie, TG, ESS, CDD, and CH) were also involved in this evaluation. The funder had no role in the study design; the study's execution, analyses, interpretation of the data; or the decision to submit the results.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Checklist for Reporting Results of Internet E-Surveys (CHERRIES).
[DOCX File , 22 KB - jmir_v24i7e34246_app1.docx]

Multimedia Appendix 2
Details on baseline measurements of demographic information and smoking behavior.
[DOCX File , 18 KB - jmir_v24i7e34246_app2.docx]

Multimedia Appendix 3
Variables included in the imputation model to test hypotheses 1a, 1b, 2a, and 2b.
[DOCX File , 13 KB - jmir_v24i7e34246_app3.docx]

Multimedia Appendix 4
Variables included in the imputation model to test hypothesis 3.
[DOCX File , 12 KB - jmir_v24i7e34246_app4.docx]

Multimedia Appendix 5
Baseline differences between participants who completed decision aid and those who did not.
[DOCX File , 19 KB - jmir_v24i7e34246_app5.docx]

Multimedia Appendix 6
Baseline differences between participants who completed time point 1 and the decision aid and those who did not complete time point 1.
[DOCX File , 19 KB - jmir_v24i7e34246_app6.docx]

Multimedia Appendix 7
Baseline differences between participants who completed time point 2 and the decision aid and those who did not complete time point 2.
[DOCX File , 19 KB - jmir_v24i7e34246_app7.docx]

Multimedia Appendix 8
Baseline differences between participants who completed time point 3 and the decision aid and those who did not complete time point 3.
[DOCX File , 19 KB - jmir_v24i7e34246_app8.docx]

Multimedia Appendix 9
Results of logistic regression for hypothesis 1b (worst-case scenario adjusted with the Revised Fagerström Test for Nicotine Dependence removed): smoking cessation after 6 months.
[DOCX File , 14 KB - jmir_v24i7e34246_app9.docx]

Multimedia Appendix 10
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2881 KB - jmir_v24i7e34246_app10.pdf]

References

1. GBD 2019 Risk Factors Collaborators. Global burden of 87 risk factors in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020 Oct 17;396(10258):1223-1249 [FREE Full text] [doi: 10.1016/S0140-6736(20)30752-2] [Medline: 33069327]
2. Bommelé J, Willemsen M. Kerncijfers roken 2019. Nationaal Expertiscentrum Tabaksontmoediging. Utrecht: Trimbos-instituut; 2020. URL: https://www.trimbos.nl/docs/452115a7-4a93-4bd4-877a-ae4a7a4d1257.pdf [accessed 2020-09-17]
3. Mujcic A, van Miltenburg C, Beenakkers E, Cruts G, Strada L, Kuin M, et al. Jaarbericht Nationale Drug Monitor 2020: Trends en cijfers over drugs, alcohol en roken. Trimbos-Instituut and Wetenschappelijk Onderzoek- en Documentatiecentrum

van het Ministerie van Justitie en Veiligheid. Utrecht: Trimbos-instituut; 2021. URL: https://www.trimbos.nl/wp-content/uploads/2021/11/AF1862-Nationale-Drug-Monitor-2020.pdf [accessed 2021-10-08]
4. Lancaster T, Stead LF. Individual behavioural counselling for smoking cessation. Cochrane Database Syst Rev 2017 Mar 31;3:CD001292 [FREE Full text] [doi: 10.1002/14651858.CD001292.pub3] [Medline: 28361496]
5. Stead LF, Perera R, Bullen C, Mant D, Hartmann-Boyce J, Cahill K, et al. Nicotine replacement therapy for smoking cessation. Cochrane Database Syst Rev 2012 Nov 14;11:CD000146. [doi: 10.1002/14651858.CD000146.pub4] [Medline: 23152200]
6. Cahill K, Stevens S, Perera R, Lancaster T. Pharmacological interventions for smoking cessation: an overview and network meta-analysis. Cochrane Database Syst Rev 2013 May 31;2013(5):CD009329 [FREE Full text] [doi: 10.1002/14651858.CD009329.pub2] [Medline: 23728690]
7. Borland R, Li L, Driezen P, Wilson N, Hammond D, Thompson ME, et al. Cessation assistance reported by smokers in 15 countries participating in the International Tobacco Control (ITC) policy evaluation surveys. Addiction 2012 Jan;107(1):197-205 [FREE Full text] [doi: 10.1111/j.1360-0443.2011.03636.x] [Medline: 21883605]
8. Bansal MA, Cummings KM, Hyland A, Giovino GA. Stop-smoking medications: who uses them, who misuses them, and who is misinformed about them? Nicotine Tob Res 2004 Dec;6 Suppl 3:S303-S310. [doi: 10.1080/14622200412331320707] [Medline: 15799593]
9. O'Connor AM. Validation of a decisional conflict scale. Med Decis Making 1995;15(1):25-30. [doi: 10.1177/0272989X9501500105] [Medline: 7898294]
10. Sun Q. Predicting downstream effects of high decisional conflict: Meta-analyses of the decisional conflict scale. University of Ottawa. 2005. URL: https://ruor.uottawa.ca/handle/10393/27050 [accessed 2021-10-08]
11. Cokkinides VE, Ward E, Jemal A, Thun MJ. Under-use of smoking-cessation treatments: results from the National Health Interview Survey, 2000. Am J Prev Med 2005 Jan;28(1):119-122. [doi: 10.1016/j.amepre.2004.09.007] [Medline: 15626567]
12. Rimal RN. Closing the knowledge-behavior gap in health promotion: the mediating role of self-efficacy. Health Commun 2000;12(3):219-237. [doi: 10.1207/S15327027HC1203_01] [Medline: 10938914]
13. Sligo FX, Jameson AM. The knowledge—behavior gap in use of health information. J Am Soc Inf Sci 2000;51(9):858-869. [doi: 10.1002/(sici)1097-4571(2000)51:9<858::aid-asi80>3.0.co;2-q]
14. Stacey D, Légaré F, Lewis K, Barry MJ, Bennett CL, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2017 Apr 12;4:CD001431 [FREE Full text] [doi: 10.1002/14651858.CD001431.pub5] [Medline: 28402085]
15. Moyo F, Archibald E, Slyer JT. Effectiveness of decision aids for smoking cessation in adults: a quantitative systematic review. JBI Database System Rev Implement Rep 2018 Sep;16(9):1791-1822. [doi: 10.11124/JBISRIR-2017-003698] [Medline: 30204670]
16. BinDhim NF, McGeechan K, Trevena L. Smartphone Smoking Cessation Application (SSC App) trial: a multicountry double-blind automated randomised controlled trial of a smoking cessation decision-aid 'app'. BMJ Open 2018 Jan 21;8(1):e017105 [FREE Full text] [doi: 10.1136/bmjopen-2017-017105] [Medline: 29358418]
17. Willemsen MC, Wiebing M, van Emst A, Zeeman G. Helping smokers to decide on the use of efficacious smoking cessation methods: a randomized controlled trial of a decision aid. Addiction 2006 Mar;101(3):441-449. [doi: 10.1111/j.1360-0443.2006.01349.x] [Medline: 16499517]
18. Kim H, Stout PA. The effects of interactivity on information processing and attitude change: implications for mental health stigma. Health Commun 2010 Mar;25(2):142-154. [doi: 10.1080/10410230903544936] [Medline: 20390680]
19. Lustria ML. Can interactivity make a difference? Effects of interactivity on the comprehension of and attitudes toward online health content. J Am Soc Inf Sci 2007 Apr;58(6):766-776. [doi: 10.1002/asi.20557]
20. Llewellyn-Thomas HA, Crump RT. Decision support for patients: values clarification and preference elicitation. Med Care Res Rev 2013 Feb;70(1 Suppl):50S-79S. [doi: 10.1177/1077558712461182] [Medline: 23124615]
21. Fagerlin A, Pignone M, Abhyankar P, Col N, Feldman-Stewart D, Gavaruzzi T, et al. Clarifying values: an updated review. BMC Med Inform Decis Mak 2013;13 Suppl 2:S8 [FREE Full text] [doi: 10.1186/1472-6947-13-S2-S8] [Medline: 24625261]
22. Witteman HO, Ndjaboue R, Vaisson G, Dansokho SC, Arnold B, Bridges JF, et al. Clarifying values: an updated and expanded systematic review and meta-analysis. Med Decis Making 2021 Oct;41(7):801-820 [FREE Full text] [doi: 10.1177/0272989X211037946] [Medline: 34565196]
23. Abhyankar P, Bekker HL, Summers BA, Velikova G. Why values elicitation techniques enable people to make informed decisions about cancer trial participation. Health Expect 2011 Mar;14 Suppl 1:20-32 [FREE Full text] [doi: 10.1111/j.1369-7625.2010.00615.x] [Medline: 20629765]
24. Feldman-Stewart D, Tong C, Siemens R, Alibhai S, Pickles T, Robinson J, et al. The impact of explicit values clarification exercises in a patient decision aid emerges after the decision is actually made: evidence from a randomized controlled trial. Med Decis Making 2012;32(4):616-626. [doi: 10.1177/0272989X11434601] [Medline: 22287534]
25. Witteman HO, Julien AS, Ndjaboue R, Exe NL, Kahn VC, Angie Fagerlin A, et al. What helps people make values-congruent medical decisions? Eleven strategies tested across 6 studies. Med Decis Making 2020 Apr;40(3):266-278. [doi: 10.1177/0272989X20904955] [Medline: 32428426]

26. Witteman HO, Gavaruzzi T, Scherer LD, Pieterse AH, Fuhrel-Forbis A, Chipenda Dansokho S, et al. Effects of design features of explicit values clarification methods: a systematic review. Med Decis Making 2016 Aug;36(6):760-776. [doi: 10.1177/0272989X16634085] [Medline: 27044883]
27. Coulter A, Stilwell D, Kryworuchko J, Mullen PD, Ng CJ, van der Weijden T. A systematic development process for patient decision aids. BMC Med Inform Decis Mak 2013;13 Suppl 2:S2 [FREE Full text] [doi: 10.1186/1472-6947-13-S2-S2] [Medline: 24625093]
28. Gültzow T, Smit ES, Hudales R, Knapen V, Rademakers J, Dirksen CD, et al. An autonomy-supportive online decision aid to assist smokers in choosing evidence-based cessation assistance: development process and protocol of a randomized controlled trial. JMIR Res Protoc 2020 Dec 15;9(12):e21772 [FREE Full text] [doi: 10.2196/21772] [Medline: 33320096]
29. Gültzow T, Smit ES, Hudales R, Dirksen CD, Hoving C. Informed decision making on the uptake of evidence-based smoking cessation assistance by individuals motivated to quit smoking: a needs assessment among potential end users and experts to inform decision aid development. Eur J Health Commun 2022 Jan 04;3(1):1-30. [doi: 10.47368/ejhc.2022.101]
30. 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]
31. Hagger MS, Hardcastle SJ, Chater A, Mallett C, Pal S, Chatzisarantis NL. Autonomous and controlled motivational regulations for multiple health-related behaviors: between- and within-participants analyses. Health Psychol Behav Med 2014 Jan 01;2(1):565-601 [FREE Full text] [doi: 10.1080/21642850.2014.912945] [Medline: 25750803]
32. Ryan RM, Patrick H, Deci EL, Williams GC. Facilitating health behaviour change and its maintenance: interventions based on Self-Determination Theory. Eur Health Psychol 2008 Mar;10(1):2-5.
33. Icon Array. Risk Science Center and Center for Bioethics and Social Sciences in Medicine. URL: http://www.iconarray.com/ [accessed 2020-01-01]
34. 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]
35. Gültzow T, Smit ES, Crutzen R, Jolani S, Hoving C, Dirksen CD. Effect study. Open Source Framework. 2021 Apr 8. URL: https://osf.io/7kcpm/ [accessed 2021-10-12]
36. Keuzehulp Visor (@keuzehulp_visor). Instagram photos and videos. URL: https://www.instagram.com/keuzehulp_visor/?hl=en [accessed 2020-03-24]
37. Gültzow T, Hoving C. Nieuw onderzoek naar keuzehulp bij stoppen-met-roken. Huisarts Wet 2021 Jan 1;64(1):10. [doi: 10.1007/s12445-020-0962-5]
38. Mudde AN, Willemsen MC, Kremers SP, de Vries H. Meetinstrumenten voor onderzoek naar stoppen met roken. The Hague, The Netherlands: Stivoro; 2000.
39. Smit ES, de Vries H, Hoving C. Effectiveness of a Web-based multiple tailored smoking cessation program: a randomized controlled trial among Dutch adult smokers. J Med Internet Res 2012 Jun 11;14(3):e82 [FREE Full text] [doi: 10.2196/jmir.1812] [Medline: 22687887]
40. Altendorf M, Hoving C, Van Weert JC, Smit ES. Effectiveness of message frame-tailoring in a Web-based smoking cessation program: randomized controlled trial. J Med Internet Res 2020 Apr 03;22(4):e17251 [FREE Full text] [doi: 10.2196/17251] [Medline: 32242826]
41. Crutzen R, Peters GJ. Scale quality: alpha is an inadequate estimate and factor-analytic evidence is needed first of all. Health Psychol Rev 2017 Sep;11(3):242-247. [doi: 10.1080/17437199.2015.1124240] [Medline: 26602990]
42. The Rosetta Stats Authors. Rosetta Stats. 2020. URL: https://sci-ops.gitlab.io/rosetta-stats/ [accessed 2021-03-25]
43. R Development Core Team. R: a language and environment for statistical computing. The R Project for Statistical Computing. 2020. URL: https://www.R-project.org/ [accessed 2020-10-12]
44. McDonald RP. Test Theory: A Unified Treatment. New York, NY, USA: Psychology Press; 2013.
45. Opleidingsniveau. Centraal Bureau voor de Statistiek. URL: https://www.cbs.nl/nl-nl/nieuws/2019/33/verschil-levensverwachting-hoog-en-laagopgeleid-groeit/opleidingsniveau [accessed 2021-04-21]
46. West R, Zatonski W, Cedzynska M, Lewandowska D, Pazik J, Aveyard P, et al. Placebo-controlled trial of cytisine for smoking cessation. N Engl J Med 2011 Sep 29;365(13):1193-1200. [doi: 10.1056/NEJMoa1102035] [Medline: 21991893]
47. Hajek P, McRobbie H, Myers K. Efficacy of cytisine in helping smokers quit: systematic review and meta-analysis. Thorax 2013 Nov;68(11):1037-1042. [doi: 10.1136/thoraxjnl-2012-203035] [Medline: 23404838]
48. Korte KJ, Capron DW, Zvolensky M, Schmidt NB. The Fagerström test for nicotine dependence: do revisions in the item scoring enhance the psychometric properties? Addict Behav 2013 Mar;38(3):1757-1763 [FREE Full text] [doi: 10.1016/j.addbeh.2012.10.013] [Medline: 23254226]
49. Module 4 : Products to Support Smoking Cessation Attempts: Working out Cigarette Equivalents. Scottish Community Pharmacy Smoking Cessation Programme. URL: http://www.smoking2.nes.scot.nhs.uk/module4/working-out-cigarette-equivalents.html [accessed 2021-05-06]
50. O'Connor AM. User Manual - Stage of Decision Making. Ottawa Hospital Research Institute. 2000. URL: https://decisionaid.ohri.ca/docs/develop/User_Manuals/UM_Stage_Decision_Making.pdf [accessed 2020-03-24]

51. O'Connor AM. User Manual - Decisional Conflict Scale. Ottawa Hospital Research Institute. 1993. URL: https://decisionaid.ohri.ca/docs/develop/User_Manuals/UM_Decisional_Conflict.pdf [accessed 2021-01-14]
52. R Studio Team. Integrated Development for R. RStudio. 2020. URL: http://www.rstudio.com [accessed 2020-10-12]
53. 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]
54. de Boer MR, Waterlander WE, Kuijper LD, Steenhuis IH, Twisk JW. Testing for baseline differences in randomized controlled trials: an unhealthy research behavior that is hard to eradicate. Int J Behav Nutr Phys Act 2015 Jan 24;12:4 [FREE Full text] [doi: 10.1186/s12966-015-0162-z] [Medline: 25616598]
55. Gruijters SL. Baseline comparisons and covariate fishing: bad statistical habits we should have broken yesterday. Eur Health Psychol 2016;18(5):205-209. [doi: 10.31234/osf.io/qftwg]
56. van Buuren S, Groothuis-Oudshoorn K, Vink G, Schouten R, Robitzsch A, Rockenschaub P, et al. mice: multivariate imputation by chained equations in R. The Comprehensive R Archive Network. 2021. URL: https://CRAN.R-project.org/package=mice [accessed 2021-03-25]
57. The CONSORT Flow Diagram. CONSORT. URL: http://www.consort-statement.org/consort-statement/flow-diagram [accessed 2021-06-09]
58. Blankers M, Smit ES, van der Pol P, de Vries H, Hoving C, van Laar M. The missing=smoking assumption: a fallacy in Internet-based smoking cessation trials? Nicotine Tob Res 2016 Jan;18(1):25-33. [doi: 10.1093/ntr/ntv055] [Medline: 25744969]
59. Jaspers MW. A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. Int J Med Inform 2009 May;78(5):340-353. [doi: 10.1016/j.ijmedinf.2008.10.002] [Medline: 19046928]
60. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018 Nov 16;20(11):e292 [FREE Full text] [doi: 10.2196/jmir.9397] [Medline: 30446482]
61. Brouwer W, Kroeze W, Crutzen R, de Nooijer J, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review. J Med Internet Res 2011 Jan 06;13(1):e2 [FREE Full text] [doi: 10.2196/jmir.1639] [Medline: 21212045]
62. Smit ES, Hoving C, Cox VC, de Vries H. Influence of recruitment strategy on the reach and effect of a Web-based multiple tailored smoking cessation intervention among Dutch adult smokers. Health Educ Res 2012 Apr;27(2):191-199. [doi: 10.1093/her/cyr099] [Medline: 22058288]
63. Bricca A, Swithenbank Z, Scott N, Treweek S, Johnston M, Black N, et al. Predictors of recruitment and retention in randomized controlled trials of behavioural smoking cessation interventions: a systematic review and meta-regression analysis. Addiction 2022 Feb;117(2):299-311. [doi: 10.1111/add.15614] [Medline: 34159677]
64. Tzelepis F, Paul CL, Williams CM, Gilligan C, Regan T, Daly J, et al. Real-time video counselling for smoking cessation. Cochrane Database Syst Rev 2019 Oct 29;2019(10):CD012659 [FREE Full text] [doi: 10.1002/14651858.CD012659.pub2] [Medline: 31684699]
65. Perski O, Crane D, Beard E, Brown J. Does the addition of a supportive chatbot promote user engagement with a smoking cessation app? An experimental study. Digit Health 2019 Sep 30;5:2055207619880676 [FREE Full text] [doi: 10.1177/2055207619880676] [Medline: 31620306]
66. Bailey RA, Pfeifer M, Shillington AC, Harshaw Q, Funnell MM, VanWingen J, et al. Effect of a patient decision aid (PDA) for type 2 diabetes on knowledge, decisional self-efficacy, and decisional conflict. BMC Health Serv Res 2016 Jan 14;16:10 [FREE Full text] [doi: 10.1186/s12913-016-1262-4] [Medline: 26762150]
67. Osaka W, Nakayama K. Effect of a decision aid with patient narratives in reducing decisional conflict in choice for surgery among early-stage breast cancer patients: a three-arm randomized controlled trial. Patient Educ Couns 2017 Mar;100(3):550-562. [doi: 10.1016/j.pec.2016.09.011] [Medline: 28277290]
68. Short CE, Smit ES, Crutzen R. Measuring psychological constructs in computer-tailored interventions: novel possibilities to reduce participant burden and increase engagement. Eur Health Psychol 2022 Jan 31;22(4):801-815.
69. Kruger J, O'Halloran A, Rosenthal AC, Babb SD, Fiore MC. Receipt of evidence-based brief cessation interventions by health professionals and use of cessation assisted treatments among current adult cigarette-only smokers: National Adult Tobacco Survey, 2009-2010. BMC Public Health 2016 Feb 11;16:141 [FREE Full text] [doi: 10.1186/s12889-016-2798-2] [Medline: 26868930]
70. Gültzow T, Hoving C, Smit ES, Bekker HL. Integrating behaviour change interventions and patient decision aids: how to accomplish synergistic effects? Patient Educ Couns 2021 Dec;104(12):3104-3108. [doi: 10.1016/j.pec.2021.04.007] [Medline: 33952401]
71. Sheridan SL, Griffith JM, Behrend L, Gizlice Z, Cai J, Pignone MP. Effect of adding a values clarification exercise to a decision aid on heart disease prevention: a randomized trial. Med Decis Making 2010;30(4):E28-E39 [FREE Full text] [doi: 10.1177/0272989X10369008] [Medline: 20484089]
72. Garvelink MM, Boland L, Klein K, Nguyen DV, Menear M, Bekker HL, et al. Decisional conflict scale findings among patients and surrogates making health decisions: part II of an anniversary review. Med Decis Making 2019 May;39(4):315-326. [doi: 10.1177/0272989X19851346] [Medline: 31142205]

73. Larsen MB, Stokholm R, Kirkegaard P, Laursen HS, Gabel P, Andersen B. Making decisions on your own: self-administered decision aids about colorectal cancer screening - a systematic review and meta-analyses. Patient Educ Couns 2022 Mar;105(3):534-546 [FREE Full text] [doi: 10.1016/j.pec.2021.07.035] [Medline: 34376303]
74. Gültzow T, Zijlstra DN, Bolman C, de Vries H, Dirksen CD, Muris JW, et al. Decision aids to facilitate decision making around behavior change in the field of health promotion: a scoping review. Patient Educ Couns 2021 Jun;104(6):1266-1285 [FREE Full text] [doi: 10.1016/j.pec.2021.01.015] [Medline: 33531158]
75. CBS. Centraal Bureau voor de Statistiek. URL: https://www.cbs.nl/ [accessed 2021-09-02]
76. Cheung KL, de Ruijter D, Hiligsmann M, Elfeddali I, Hoving C, Evers SM, et al. Exploring consensus on how to measure smoking cessation. A Delphi study. BMC Public Health 2017 Nov 21;17(1):890 [FREE Full text] [doi: 10.1186/s12889-017-4902-7] [Medline: 29162043]
77. West R, Hajek P, Stead L, Stapleton J. Outcome criteria in smoking cessation trials: proposal for a common standard. Addiction 2005 Mar;100(3):299-303. [doi: 10.1111/j.1360-0443.2004.00995.x] [Medline: 15733243]
78. Hughes JR, Keely J, Naud S. Shape of the relapse curve and long-term abstinence among untreated smokers. Addiction 2004 Jan;99(1):29-38. [doi: 10.1111/j.1360-0443.2004.00540.x] [Medline: 14678060]
79. Patrick DL, Cheadle A, Thompson DC, Diehr P, Koepsell T, Kinne S. The validity of self-reported smoking: a review and meta-analysis. Am J Public Health 1994 Jul;84(7):1086-1093. [doi: 10.2105/ajph.84.7.1086] [Medline: 8017530]
80. Internet access and use statistics—households and individuals. Eurostat. 2016. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Archive:Internet_access_and_use_statistics_-_households_and_individuals&oldid=379591 [accessed 2020-11-28]

Abbreviations

CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
DA: decision aid
FTND-R: Revised Fagerström Test for Nicotine Dependence
IPDAS: International Patient Decision Aid Standards Collaboration
MI: multiple imputation
OR: odds ratio
OSF: open science framework
RCT: randomized controlled trial
SDT: self-determination theory
t=1: time point 1
t=2: time point 2
t=3: time point 3
VCM: value clarification method

Edited by A Mavragani; submitted 13.10.21; peer-reviewed by K Selby, P van der Keylen, R Plackett, M Kopka, H Ayatollahi; comments to author 29.12.21; revised version received 17.03.22; accepted 07.04.22; published 15.07.22.

Please cite as:
Gültzow T, Smit ES, Crutzen R, Jolani S, Hoving C, Dirksen CD
Effects of an Explicit Value Clarification Method With Computer-Tailored Advice on the Effectiveness of a Web-Based Smoking Cessation Decision Aid: Findings From a Randomized Controlled Trial
J Med Internet Res 2022;24(7):e34246
URL: https://www.jmir.org/2022/7/e34246
doi:10.2196/34246
PMID:35838773

Original Paper

Patients' Willingness and Ability to Identify and Respond to Errors in Their Personal Health Records: Mixed Methods Analysis of Cross-sectional Survey Data

Rachael Lear[1], RN, BA, PhD; Lisa Freise[1], BSc, MSc; Matthew Kybert[2], BEng; Ara Darzi[1], PC, KBE, MD; Ana Luisa Neves[1], MSc, MD, PhD; Erik K Mayer[1], BSc, MBBS, MRCS, PhD

[1]National Institute for Health Research Imperial Patient Safety Translational Research Centre, Institute of Global Health Innovation, Imperial College London, London, United Kingdom

[2]Imperial College Healthcare National Health Service Trust, London, United Kingdom

Corresponding Author:
Rachael Lear, RN, BA, PhD
National Institute for Health Research Imperial Patient Safety Translational Research Centre
Institute of Global Health Innovation
Imperial College London
St Mary's Campus 10F Queen Elizabeth Queen Mother Wing
Praed Street
London, W2 1NY
United Kingdom
Phone: 44 020 7589 5111
Email: r.lear12@imperial.ac.uk

Abstract

Background: Errors in electronic health records are known to contribute to patient safety incidents; however, systems for checking the accuracy of patient records are almost nonexistent. Personal health records (PHRs) enabling patient access to and interaction with the clinical records offer a valuable opportunity for patients to actively participate in error surveillance.

Objective: This study aims to evaluate patients' willingness and ability to identify and respond to errors in their PHRs.

Methods: A cross-sectional survey was conducted using a web-based questionnaire. Patient sociodemographic data were collected, including age, sex, ethnicity, educational level, health status, geographical location, motivation to self-manage, and digital health literacy (measured using the eHealth Literacy Scale tool). Patients with experience of using the Care Information Exchange (CIE) portal, who specified both age and sex, were included in these analyses. The patients' responses to 4 relevant survey items (closed-ended questions, some with space for free-text comments) were examined to understand their willingness and ability to identify and respond to errors in their PHRs. Multinomial logistic regression was used to identify patients' characteristics that predict the ability to understand information in the CIE and willingness to respond to errors in their records. The framework method was used to derive themes from patients' free-text responses.

Results: Of 445 patients, 181 (40.7%) "definitely" understood the CIE information and approximately half (220/445, 49.4%) understood the CIE information "to some extent." Patients with high digital health literacy (eHealth Literacy Scale score ≥26) were more confident in their ability to understand their records compared with patients with low digital health literacy (odds ratio [OR] 7.85, 95% CI 3.04-20.29; *P*<.001). Information-related barriers (medical terminology and lack of medical guidance or contextual information) and system-related barriers (functionality or usability and information communicated or displayed poorly) were described. Of 445 patients, 79 (17.8%) had noticed errors in their PHRs, which were related to patient demographic details, diagnoses, medical history, results, medications, letters or correspondence, and appointments. Most patients (272/445, 61.1%) wanted to be able to flag up errors to their health professionals for correction; 20.4% (91/445) of the patients were willing to correct errors themselves. Native English speakers were more likely to be willing to flag up errors to health professionals (OR 3.45, 95% CI 1.11-10.78; *P*=.03) or correct errors themselves (OR 5.65, 95% CI 1.33-24.03; *P*=.02).

Conclusions: A large proportion of patients were able and willing to identify and respond to errors in their PHRs. However, some barriers persist that disproportionately affect the underserved groups. Further development of PHR systems, including incorporating channels for patient feedback on the accuracy of their records, should address the needs of nonnative English speakers and patients with lower digital health literacy.

(J Med Internet Res 2022;24(7):e37226) doi:10.2196/37226

KEYWORDS

electronic health records; personal health records; patient participation; errors; patient safety; digital health literacy

Introduction

Background

Errors in electronic health records (EHRs) are not uncommon and are known to contribute to patient safety incidents [1]. Up to 60% of patient records may include inaccuracies or omissions, such as errors in patients' diagnoses, medical history, medications, allergies, test results, procedures, contact information, and appointment details [2-6]. Early EHR systems largely reflect traditional paper–based records with records containing large amounts of unstructured free text stored in a digital format. As such, it can be time-consuming for clinicians to locate pertinent information in the patient's record, and data entry errors can occur when health care professionals copy and paste outdated information from other parts of the record [7,8]. Computerized physician order entry and clinical decision support systems were introduced into EHRs to reduce error, and they have undoubtedly benefited clinicians and patients [9]. However, the unintended consequences of poorly designed EHR functionality and associated usability issues (eg, alert fatigue) have compounded the problem of EHR errors, creating new threats to patient safety [8,10,11]. In addition, lack of EHR interoperability can mean that important medical information entered by clinicians in different organizations is absent from the records [12]. As health care professionals use EHR data to inform clinical decision-making, errors and omissions may lead to delayed diagnoses, inappropriate treatment, and medication safety incidents [1,13,14]. Despite these known problems, there are no routine mechanisms for checking the accuracy of patients' electronic records [4].

Whereas an EHR is a computer record that originates with and is controlled by health care professionals, a personal health record (PHR) can be generated by clinicians but is controlled by the patient. PHRs not only provide patients with access to their clinical records but also enable them to input their own data and manage who sees their records among different providers [15,16]. Recent widespread adoption of PHR systems has been driven by emphasis on transparency, an appetite for patients to become partners in their own care, and the need for health record integration not currently provided by EHR systems [15,17,18]. Evidence suggests that sharing electronic records with patients positively affects several domains of care quality, including enhanced patient engagement and involvement of informal caregivers, adherence to treatment, and timely follow-up [6,14,19-22]. Despite concerns that PHR systems could damage the patient-physician relationship if patients were to identify errors in their records, research indicates that sharing records with patients improves trust in care providers, and patients feel empowered by the opportunity to check their records for accuracy [20,22]. As the trend toward patients being able to access their electronic records accelerates through rapid digital transformation of health care systems [17], there is a valuable opportunity for patients to play a role in identifying and addressing erroneous information in the PHR. Preliminary evidence suggests that most patients can understand the information in their records to identify potential errors [2]; however, further research is needed to explore how patient involvement could be leveraged to improve the accuracy of health records. A better understanding of patients' views around PHR error surveillance is important to uncover factors that may exacerbate the *digital divide* and linked health inequalities, such that future initiatives to involve patients in addressing PHR error are accessible to diverse patient groups.

Objectives

The aim of this study was to evaluate patients' willingness and ability to identify and respond to PHR errors to inform future error surveillance initiatives.

Methods

Study Design, Participants, and Data Collection

A cross-sectional study, using a web-based survey, explored patients' views and experiences of using the Care Information Exchange (CIE), the largest shared PHR program in the United Kingdom. In 2018, the CIE was rolled out to the diverse 2.3 million patients treated in North West London. At the time of this survey, the CIE held patient information from both hospitals and general practitioner practices in North West London, and records from 15 hospitals outside London—in Birmingham, Bristol, Liverpool, Manchester, Scotland, and Wales [23].

The questionnaire was open for completion between July 1, 2018, and July 1, 2019. All patients registered with the CIE during the study period (N=27,411) were eligible to participate. The survey was administered via Qualtrics (a web-based survey platform). Patients registered with the CIE were invited by email to complete the questionnaire via a weblink in the portal. The email explained the purpose of the study, and informed consent was obtained. Patients had to be aged at least 18 years to register with the CIE. Not all patients registered with the CIE had used the portal; for the analyses presented, we only included respondents who indicated that they had used the CIE. We excluded patients who did not provide basic demographics regarding age and sex. Considering this population, a CI of 95%, and a margin of error of 5%, the minimum sample size to ensure representativeness was calculated as 379 respondents.

We have previously characterized individuals registered with the CIE and evaluated the differences between users and nonusers of the CIE with respect to their sociodemographic characteristics, health status, and motivation related to being involved in their own care [23]. Our findings highlight the importance of addressing educational aspects (educational level and digital literacy) to ensure equitable and sustainable portal adoption [23]. Building on this previous work, we sought to understand how portal use could be leveraged to improve patient safety and care quality. Patient portals offer users the

opportunity to actively contribute to patient safety by identifying errors in PHRs and taking action to ensure that these are rectified. To conduct an initial assessment of the acceptability and feasibility of this patient safety strategy, this study aimed to evaluate the CIE users' willingness and ability to identify and respond to errors in their EHRs.

Patients' responses to 4 specific questions were analyzed (Textbox 1). The questions were multiple choice and closed ended, with some responses prompting the respondent to elaborate using free text.

The following sociodemographic information was used in this analysis to identify predictors of patients' willingness and ability to identify and respond to errors in their records: age, sex, ethnicity, native language, education level, geographical location, and health status. Respondents' level of motivation to be involved in their own care was assessed via a multiple choice question ("In general, how motivated to be involved in your healthcare are you?" Possible responses: "A little," "A moderate amount," "A lot," or "Very much"). In addition, digital health literacy was assessed using the eHealth Literacy Scale (eHEALS), developed and validated by Norman and Skinner [24]. The eHEALS tool is an 8-item measure of patients' combined knowledge, comfort, and perceived skills in finding, evaluating, and using internet health resources for health problems [24]. The 8 items are answered on a 5-point Likert scale (1-strongly disagree to 5-strongly agree); total eHEALS scores range from 8 to 40, with a higher score indicating higher digital literacy.

Textbox 1. Questionnaire items and format of responses.

Questionnaire items and responses

1. Did you understand the information that you saw on Care Information Exchange? (*possible responses*):

- Yes, definitely
- Yes, to some extent
- No, please specify <*free-text response*>
- Not sure, please specify <*free-text response*>

2. When using Care Information Exchange, did you notice any errors in your record? (*possible responses*):

- No, I did not notice any errors
- Yes, I did notice errors. Please specify <*free-text response*>

3. If you were to see an error in your medical record, what would you like to be able to do? (*possible responses*):

- Nothing
- Flag it up to my health care professional for correction (to promote patient safety, the questionnaire included the following statement: "If you have noticed errors in your record and have not already reported them, then it is important to contact your clinical team").
- Correct it myself
- Unsure

4. Which types of errors would you personally feel comfortable correcting? (*tick all that apply*):

- Personal information
- Medication names
- Medication doses
- Physicians' notes
- Dates of appointments
- Diagnosis
- Other, please specify <*free-text response*>

Analysis

To assess the effects of excluding patients with missing data regarding age and sex (78/523, 14.9%), we ran a Pearson chi-square test of homogeneity (c^2) to compare the distribution of responses to survey items between the analysis sample and the missing data sample. We used descriptive statistics to quantitatively summarize respondent characteristics and users' responses to structured survey items. Counts and proportions were calculated for categorical variables; means and SDs were calculated for continuous variables. Using the first part of the respondents' postcodes, geographical location was categorized according to London's official postal districts, with an additional category "other" for patients who reside outside London's postal districts. Age was categorized into bands (<30, 31-40, 41-50, 51-65, ≥65), ethnicity was categorized into White, and all other ethnic groups were combined owing to the small numbers of

patients self-identifying to individual categories of ethnic minority background.

Multinomial Regression Analysis

We conducted multinomial regression analyses to identify the characteristics of patients that predicted (1) their ability to identify errors in their records (ie, to what extent did they understand their records?) and (2) their willingness to respond to errors in their records (ie, what action would they like to be able to take if they noticed an error in their records?) To facilitate regression analyses in the context of sparse counts (Multimedia Appendices 1 and 2), *age*, *motivation to be involved in own care*, and *digital health literacy* were treated as dichotomous variables and respondents reporting sex as "other" were excluded. After a relevant literature review, we selected an eHEALS score ≥26 to indicate higher digital health literacy and <26 to indicate lower digital health literacy [25-29]. Univariate multinomial logistic analyses were initially performed to identify potential predictors to include in the multivariable model. As suggested by Hosmer et al [30,31], we adopted the following approach to variable selection: (1) variables that demonstrated significance (P<.25) in the univariate analyses were entered into the preliminary multivariable model; (2) variables that were nonsignificant at P>.05 according to the likelihood ratio test were then removed one at a time according to the variable with the highest P value (backward elimination); and (3) to check for suppressor effects, variables excluded during backward selection were re-entered separately into the regression model (forward selection). Only variables that were significant at P<.05 (likelihood ratio test) were retained in the final multinomial regression models. Model quality comparisons were conducted using the Akaike Information Criterion [32], and the goodness of fit was assessed using Pearson chi-square statistic [31]. Effect estimates are presented as odds ratios (ORs) with 95% CIs. Analyses were conducted using Microsoft Excel (version 16.54; Microsoft Corporation) and SPSS (version 27; IBM Corp).

Framework Analysis of Free-Text Responses

Unstructured, free-text responses were analyzed to identify emerging themes using the framework analysis method described by Ritchie and Spencer [33]. Framework analysis is a transparent and systematic approach to qualitative analysis that enables researchers to interpret data through a 5-step process: (1) familiarization with the data; (2) identification of a thematic framework (themes may be identified a priori or emerge from the data itself); (3) indexing, to explore the fit of the theoretical framework to the data; (4) charting, which involves summarizing the data into theoretical charts; and (5) mapping and interpretation, which involves checking or reviewing and synthesizing the data set as a whole [33]. In all, 2 coders (RL and LF), both experienced with framework analysis, independently coded the data. The coders discussed differences in coding to reach a consensus. Organized frameworks of inductively and deductively derived themes and subthemes were generated and applied across free-text responses. The coders worked together to complete the other stages of the analysis. Themes and subthemes were presented with verbatim quotes from the CIE users' free-text responses.

Ethics Approval

The study was approved as a Service Evaluation at Imperial College Healthcare National Health Service Trust (registration number: 296/2018).

Reporting

We followed the reporting recommendations in the Strengthening the Reporting of Observational Studies in Epidemiology statement (Multimedia Appendix 3) [34].

Results

Respondent Characteristics

Of 1083 patients who responded to the survey, 674 (62.23%) patients indicated that they had previously used the CIE, and of these *CIE users*, 523 (77.6%) went on to complete some or all the questionnaire. The CIE users who provided basic demographic details regarding both age and sex were included in the analysis (445/523, 85.1%; +117% of the minimum target sample size); 14.9% (78/523) of respondents with missing data for either age or sex were excluded.

Of the 445 respondents, 276 (62%) were women and most were aged ≥51 years (313/445, 70.3%). Approximately two-thirds (284/445, 63.8%) of the respondents resided in London and a further 32.6% (145/445) of the respondents lived in other geographical locations in England. More than 1 in 5 (97/445, 21.8%) participants belonged to an ethnic minority group. Most were educated to a degree level or higher (292/445, 65.6%), and the mean digital literacy (eHEALS) score was 33.6 (SD 6.4). Most patients (278/445, 62.5%) considered themselves to be highly motivated in their own care. Approximately one-third (162/445, 36.4%) of the patients reported poor health status, whereas 39.8% (177/445) reported being in good health. Most patients (284/445, 63.8%) reported using the CIE at least once a month. Patient characteristics are shown in Table 1.

Table 1. Respondent characteristics (N=445).

Characteristics	Respondents
Sex, n (%)	
Male	167 (37.5)
Female	276 (62)
Other	2 (0.4)
No response	N/A[a]
Age group (years), n (%)	
<30	22 (4.9)
31 to 40	48 (10.8)
41 to 50	62 (13.9)
51 to 64	166 (37.3)
>65	147 (33)
No response	N/A
Ethnicity, n (%)	
Ethnic minority	97 (21.8)
White	343 (77.1)
No response	5 (1.1)
Geographic location, n (%)	
London	284 (63.8)
Other location in England	145 (32.6)
No response	16 (3.6)
Education, n (%)	
Secondary school or below	118 (26.5)
Undergraduate or professional degree	180 (40.4)
Postgraduate or higher	112 (25.2)
No response	35 (7.9)
Language, n (%)	
English	379 (85.2)
Non-English	58 (13)
No response	8 (1.8)
eHealth literacy (eHEALS[b] score), mean (SD; range)	33.6 (6.4; 8-40)
Overall health status, n (%)	
Good or very good	177 (39.8)
Neither good nor poor	106 (23.8)
Poor or very poor	162 (36.4)
No response	0 (0)
Motivation to be involved in own care, n (%)	
Not very much	6 (1.3)
A moderate amount	43 (9.7)
A lot	116 (26.1)
Very much	278 (62.5)
No response	2 (0.4)

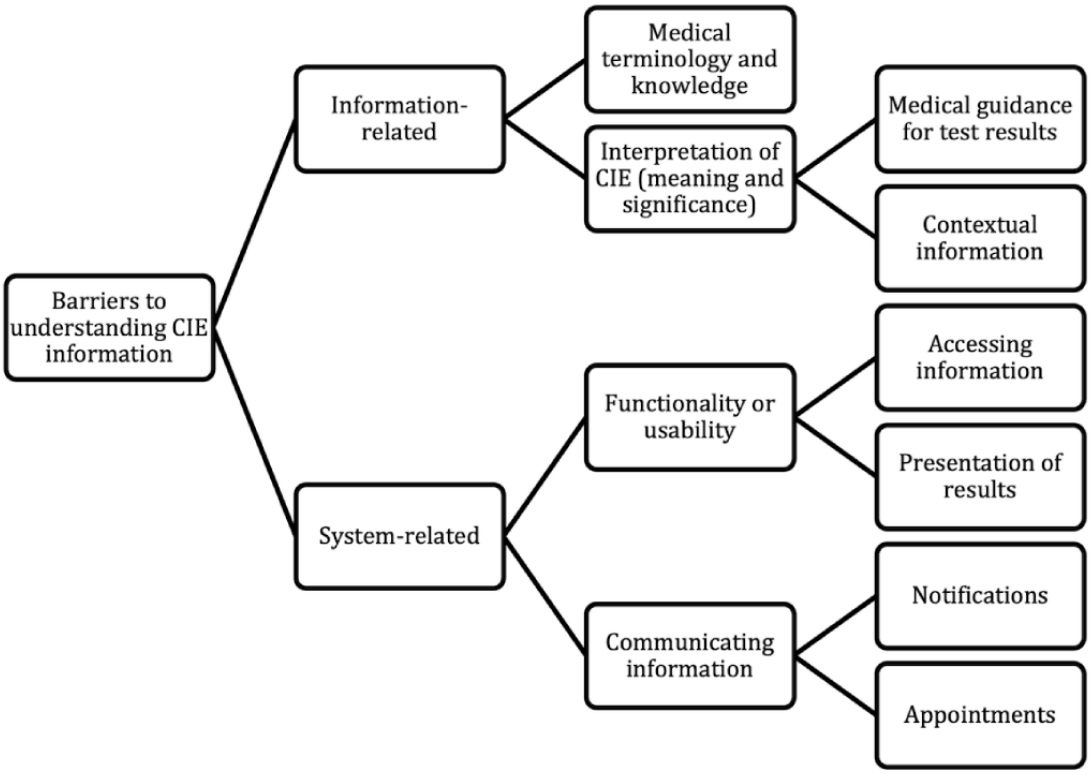
[a]N/A: not applicable.

[b]eHEALS: eHealth Literacy Scale.

Patients' Understanding of Information in the CIE

Of the 445 respondents, 181 (40.7%) reported that they "definitely" understood the CIE information and 220 (49.4%) reported understanding the CIE information "to some extent." Few patients answered "No" (31/445, 7%) or "Not sure" (13/445, 2.9%); these patients were asked to elaborate on their responses and 93% (41/44) of patients provided free-text comments. Through the mapping and interpretation of coded responses, two primary categories of barriers to understanding the CIE information were identified: (1) information-related barriers (ie, users have difficulty understanding the information itself) and (2) system-related barriers (ie, problems related to the CIE system impede users' understanding of information). These categories consist of several themes and subthemes as outlined in Figure 1.

Figure 1. Thematic map of barriers to understanding the Care Information Exchange (CIE) information.



Regarding *information-related barriers*, medical terminology or users' lack of medical knowledge were identified as barriers to understanding the CIE information as articulated by 2 patients:

> *...when it is hugely medical, as one would expect for Doctors records, then I don't always understand what I am looking at, or the relevance of its meaning.* [patient ID 280]

> *It's great having access to blood results and radiology reports but the jargon. I don't understand much of it.* [patient ID 376]

Patients often struggled to interpret the meaning or significance of information on the CIE, particularly in relation to the interpretation of test results ("Didn't know whether a test result mattered or not" [patient ID 282]). Some patients would like more medical guidance to help with understanding the CIE information, as articulated by a patient:

> *There is no room or option for doctor comments so I do not understand my test results and this makes me worried when I can see them but do not understand if they are ok or not!* [patient ID 412]

For other patients, additional contextual information would have assisted with understanding information ("Blood tests results out of range and not knowing context" [patient ID 363]). In terms of *system-related barriers* to understanding the CIE information, patients described problems with the way in which test results were presented, including how results are displayed ("Results are virtually unreadable" [patient ID 57]) or how they are listed in the CIE:

> *[...] difficult to find some test results as they are listed under the medical name and the patient may not recognize the test under these names.* [patient ID 195]

Being able to access information was problematic for some patients; for example, a patient (patient ID 153) pointed out that that the CIE system is not accessible to screen reader users; another patient reported that "Information is too difficult to navigate when loaded separately" (patient ID 235). Some patients reported that the CIE notifications were confusing ("All these messages and they were not clear what they were for"

[patient ID 428]), whereas others reported problems with how appointment information is communicated on the CIE, such as the following:

> *Often no detail given so it says I have an appointment but I don't know who with or why.* [patient ID 302]

Overall, patients' free-text responses provided fewer examples of system-related barriers than information-related barriers and a significant proportion of responses related to difficulties in understanding the test results.

Patient Characteristics Associated With Understanding Information on the CIE

Patient characteristics and responses to the question "Did you understand the information you saw on CIE?" were entered into univariate and multivariable multinomial regression models to identify patient characteristics that predicted the perceived ability to understand information in the electronic record. The final multivariable multinomial regression model with 2 predictor variables (digital health literacy and motivation) predicted significantly better than the null (intercept) model (P<.001), and Pearson chi-square statistic suggested that the model fit the data well (χ^2_2=2.4; P=.30). In the final multivariable model, digital health literacy was the only variable independently associated with understanding the CIE information (Table 2). Patients with higher digital health literacy (eHEALS score ≥26) were 6 times more likely to report "definitely" understanding the CIE information, compared with patients reporting not being able to understand the CIE information (OR 6.07, 95% CI 1.70-21.57; P=.005). Sensitivity analyses assessing the effects of including or excluding predictor variables that demonstrated significance in the univariate analyses did not alter this result.

Table 2. Multinomial regression results of users' sociodemographic characteristics and perceived level of understanding of information in the Care Information Exchange.

	Yes, to some extent vs no		Yes, definitely vs no		Yes, to some extent vs no		Yes, definitely vs no	
	Univariate odds ratio (95% CI)	*P* value	Univariate odds ratio (95% CI)	*P* value	Multivariable odds ratio (95% CI)	*P* value	Multivariable odds ratio (95% CI)	*P* value
Sex								
Female	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Male	0.76 (0.36-1.61)	.47	0.49 (0.23-1.06)	.07	__[a]	—	—	—
Age (years)								
≥65	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
≤64	1.33 (0.60-2.96)	.48	1.58 (0.70-3.57)	.27	N/A[b]	N/A	N/A	N/A[b]
Ethnicity								
White	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Ethnic minority	2.48 (0.72-8.53)	.15	3.18 (0.92-10.97)	.07	—	—	—	—
Native language								
Non-English	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
English	0.41 (0.09-1.78)	.23	0.45 (0.10-2.00)	.29	—	—	—	—
Education								
Secondary or below	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Undergraduate or professional	0.47 (0.16-1.82)	.18	0.66 (0.22-1.96)	.45	—	—	—	—
Postgraduate or higher	0.33 (0.12-1.04)	.06	0.58 (0.18-1.82)	.35	—	—	—	—
Digital literacy								
Lower digital health literacy	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Higher digital health literacy	1.47 (0.51-4.18)	.48	6.23 (1.76-22.06)	.005	1.51 (0.53-4.34)	.44	6.07 (1.70-21.57)	.005
Health status								
Neither good nor poor	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Poor	1.48 (0.51-4.29)	.48	2.58 (0.86-7.75)	.09	N/A	N/A	N/A	N/A
Good	0.63 (0.25-1.57)	.32	1.29 (0.86-7.75)	.60	N/A	N/A	N/A	N/A
Motivation to be involved in own care								
Not very much or a moderate amount	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference

	Yes, to some extent vs no		Yes, definitely vs no		Yes, to some extent vs no		Yes, definitely vs no	
	Univariate odds ratio (95% CI)	*P* value	Univariate odds ratio (95% CI)	*P* value	Multivariable odds ratio (95% CI)	*P* value	Multivariable odds ratio (95% CI)	*P* value
A lot or very much	1.08 (0.39-3.01)	.89	3.29 (1.04-10.39)	.04	0.93 (0.30-2.88)	.90	2.99 (0.85-10.61)	.09

[a]Variable excluded from the final multivariable model using a backward elimination approach.

[b]N/A: not applicable (variable excluded from the multivariable analyses because of nonsignificance [*P*>.25] in the univariate analyses).

Errors That Patients Have Noticed in Their Records

Nearly 1 in 5 patients (79/445, 17.8%) reported that they had noticed errors in their medical records. In all, 97% (77/79) of the patients provided further information regarding the nature of these errors. The themes and subthemes that emerged from the analysis of these free-text responses together with illustrative quotes are presented in Table 3. The most prominent theme emerging from patients' free-text responses was *incorrect information*. This theme described 6 categories of incorrect information: patients' details, appointments, medical history or diagnoses, measurements or results, medications, and letters or correspondence. A smaller number of respondents described information that was either incomplete or missing from the record entirely. A few patients noted instances where results or letters were missing from the record; however, most patients noted instances of missing or incomplete information related to appointments. Conflicting appointment information also emerged as a distinct theme; several users described instances of the CIE appointment information that did not match information communicated via other means (eg, through phone calls or letters). There appears to be a general lack of trust in appointment information listed on the CIE as articulated by a user:

> *I don't know. Hard to confirm whether an appointment is real or a mistake as I have not received a letter or email notification to confirm appointment [sic] listed.* [patient ID 303]

The final theme *information belonging to a different patient* contained only 2 responses: a CIE user had noticed "another patient's clinic letter" (patient ID 164) in their record, whereas another respondent reported that information belonging to a different patient had appeared in her midwifery notes.

Table 3. Types of errors patients have noticed in their records: themes, subthemes, and illustrative quotes.

Themes and subthemes	Illustrative quotes
Incorrect information	
Patient details	"incorrect NHS[a] number" (patient ID 77); "My address is wrong" (patient ID 79).
Appointments	"Appointment times are incorrect" (patient ID 117); "On one occasion, the system indicated that I had missed an appointment when I was definitely there" (patient ID 178).
Medical history or diagnoses	"Diagnosis of cancer which I do not have" (patient ID 317); "my GP[b] said, 'he became depressed.' No, I didn't" (patient ID 120).
Measurements or results	"A time on a test was wrong" (patient ID 304); "MRSA result appeared on my record for a test I never took" (patient ID 95).
Medications	"An error on the dosage of one of my medications" (patient ID 280); "I have never received ibuprofen on prescription" (patient ID 128).
Letters or correspondence	"Incorrect information on discharge notice from A&E[c]" (patient ID 173); "Errors in my consultants [sic] letter" (patient ID 271).
Missing or incomplete information	
Appointments	"Some appointments are not shown" (patient ID 406).
Measurements or results	"Blood test and urine results are missing" (patient ID 413).
Letters or correspondence	"2 notifications disappeared" (patient ID 264).
Conflicting appointment information	"I am shown as having multiple appointments at different dates at the same clinic. The last time I turned up for one of these I was told there was no appointment [...] very confusing!" (patient ID 226).
Information belonging to a different patient	"Another patient's clinic letter—major breach of confidence" (patient ID 164).

[a]NHS: National Health Service.

[b]GP: general practitioner.

[c]A&E: accident and emergency.

Responding to Errors in the Medical Record

When asked to consider how they would like to respond to errors in their records, most patients (272/445, 61.1%) would like to flag up errors to their health professionals for correction. Although some patients (57/445, 12.8%) were unsure what action they would like to take, only a small proportion (16/445, 3.6%) said they would not take any action. Approximately one-fifth (91/445, 20.4%) of the respondents were willing to correct errors themselves. Of the 91 CIE patients who were willing to correct errors themselves, most were comfortable correcting errors in their personal information (88/91, 97%), and approximately two-thirds were willing to correct medication doses (58/91, 64%) or medication names (56/91, 62%). A smaller proportion reported willingness to correct physicians' notes (33/91, 36%) or diagnoses (31/91, 34%).

Patient Characteristics Associated With Willingness to Correct Errors

Patient characteristics and responses to the question, "If you were to see an error in your medical record, what would you like to be able to do?" were entered into univariate and multivariable multinomial regression models to identify patient characteristics that predicted willingness to take action to address errors in the record. The reference category for these analyses was *do nothing*. The results are presented in Table 4. The final multivariable multinomial regression model with 2 predictor variables (language and health status) predicted significantly better than the null (intercept) model (*P*=.04) and the model fit the data well (χ^2_4=6.7; *P*=.16). In multivariable analyses, users' native language and health status predicted their willingness to take action to address errors in the record. Native English speakers were more likely to select *flag up errors to health professionals* (OR 3.45, 95% CI 1.11-10.78; *P*=.03). Native English speakers were also more likely to be willing to correct errors in their medical records themselves (OR 5.65, 95% CI 1.33-24.03; *P*=.02). Compared with patients reporting that their health was neither good nor poor (reference group), patients reporting good health status were more likely to correct errors themselves (OR 3.84, 95% CI 1.07-13.75; *P*=.04). However, there was no convincing evidence that poor health status increased the odds of being willing to correct errors (OR 3.35, 95% CI 0.89-12.61; *P*=.08). Post hoc sensitivity analyses did not affect any of these findings.

Table 4. Multinomial regression results of users' sociodemographic characteristics and willingness to flag up or correct errors in the Care Information Exchange.

Characteristics	Flag it up[a] vs do nothing		Correct it myself vs do nothing		Flag it up[a] vs do nothing		Correct it myself vs do nothing	
	Univariate odds ratio (95% CI)	*P* value	Univariate odds ratio (95% CI)	*P* value	Multivariable odds ratio (95% CI)	*P* value	Multivariable odds ratio (95% CI)	*P* value
Sex								
Female	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Male	0.71 (0.25-2.02)	.52	0.59 (0.20-1.78)	.35	N/A[b]	N/A	N/A	N/A
Age (years)								
≥65	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
≤64	0.81 (0.25-2.64)	.72	0.73 (0.21-2.54)	.62	N/A	N/A	N/A	N/A
Ethnicity								
White	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Ethnic minority	0.48 (0.16-1.50)	.21	0.48 (0.14-1.58)	.23	—[c]	—	—	—
Native language								
Non-English	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
English	3.13 (1.02-9.56)	.05	3.59, (0.51-2.69)	.04	3.45 (1.11-10.78)	.03	5.65 (1.33-24.03)	.02
Education								
Secondary or below	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Undergraduate or professional	1.57 (0.44-5.6)	.49	1.36 (0.36-5.21)	.65	N/A	N/A	N/A	N/A
Postgraduate or higher	1.55 (0.36-6.75)	.56	1.80 (0.39-8.32)	.45	N/A	N/A	N/A	N/A
Digital health literacy								
Low digital health literacy	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
High digital health literacy	2.25 (0.47-10.83)	.31	144 (0.28-7.51)	.66	N/A	N/A	N/A	N/A
Health status								
Neither good nor poor	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Poor	1.45 (0.45-4.68)	.54	3.73 (1.00-13.85)	.05	1.34 (0.41-4.41)	.63	3.35 (0.89-12.61)	.08
Good	2.61 (0.71-9.58)	.15	5.32 (1.27-22.25)	.02	2.82 (0.76-10.54)	.12	3.84 (1.07-13.75)	.04
Motivation to be involved in own care								
Not very much or a moderate amount	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
A lot or very much	2.09 (0.56-7.78)	.27	2.37 (0.56-10.09)	.25	—	—	—	—

[a]Flag it up to my health care professional for correction.

[b]N/A: not applicable (variable excluded from the multivariable analyses because of nonsignificance [*P*<.25] in univariate analyses).

[c]Variable excluded from the final multivariable model using a backward elimination approach.

Missing Data Analysis

Meaningful comparisons of sociodemographic characteristics between the analysis sample and the missing data sample were not possible because of considerable missing data in the group of 14.9% (78/523) of respondents excluded from this analysis (Multimedia Appendix 4). There were no differences in the distribution of structured questionnaire responses between the analysis sample and the missing data sample (Multimedia Appendix 5).

Discussion

Summary of Key Findings

A large proportion of patients could understand health information held in the CIE (fully or to some extent), and some reported that they had previously noticed errors in their records across a range of categories, including patient details, diagnoses, medical history, results, and medications. They also highlighted errors in letters or correspondence and in appointment information. Nearly two-thirds (272/445, 61.1%) of patients would like to be able to flag up PHR errors to health professionals, and some were willing to correct errors themselves; however, these patients were more likely to be native English speakers. A minority of patients with low digital health literacy had difficulty understanding information in their records. Barriers included a lack of medical knowledge or guidance, issues with portal functionality and usability, and inadequate presentation or communication of information in the portal.

Comparison With Previous Literature

Despite policy and provider commitments enabling patient access to digital health records [17,35,36], mechanisms for systematically checking the accuracy of patient health data are almost nonexistent. Although evidence for patient involvement in improving the accuracy of their health records is limited, the findings of our study are consistent with previous research suggesting that many patients are willing and able to identify errors in their records, and their involvement may help reduce medication errors, diagnostic and treatment delays, and wasteful duplication of tests or procedures [2,4,37]. Patients can identify errors and omissions across a range of categories, including current and past diagnoses, medical or social histories, medications and allergies, procedures, test results, and appointment scheduling, with many patient-reported errors having the potential to affect care [4,7,37,38]. Evidence suggests that in many cases, patients can identify serious mistakes with clinically relevant implications that might otherwise go undetected [4,6,37-39]. This highlights the potential safety gains that can be achieved by introducing mechanisms empowering patients to engage in improving the accuracy of the personal health information. A novel finding of this study is that approximately 1 in 5 (91/445, 20.4%) patients are willing to correct errors themselves; however, no previous studies have evaluated the feasibility of this approach; key questions include which categories of errors patients could reasonably be expected to correct, as well as the safety of introducing such functionality into PHR systems.

Although patients are interested in improving the accuracy of their PHRs, certain sociodemographic factors appear to predict their readiness to participate in error surveillance. We found a significant association between native language and willingness to either correct errors or flag them up to health professionals. Similar findings from a large US study demonstrated that patients who speak a language other than English or Spanish as their primary language are less likely to report serious mistakes in electronic ambulatory visit notes [4]. We did not find any significant associations among age, sex, ethnicity, or educational level, and willingness to act on errors in the records. However, other studies have reported that patients who are male, younger, less educated, and those self-identifying as Black or African American, Asian, or from mixed ethnic backgrounds are less likely to report mistakes in their electronic records [4,38]. These findings emphasize that issues of equity must be considered when designing patient-facing error surveillance systems, such that minority groups and patients who choose not to participate in addressing error accuracy are not disadvantaged by a *digital divide* [40].

Equitable patient involvement in error surveillance is, on a basic level, contingent on all patients being able to access and understand PHR information. The characteristics of our sample are consistent with those of previous studies, reinforcing a distinct demographic profile associated with portal use, with users tending to be female, older, highly educated, with higher digital health literacy [14,22,23,41]. Previous studies have demonstrated that patients who struggle to use web-based information for health purposes are less likely to use PHR systems and that lacking these eHealth literacy skills may contribute to adverse health outcomes [23,42-44]. Our analysis reveals that patients with lower digital health literacy are also less likely to understand the information held in the PHR, which may partly explain the health disparities observed in this patient group. Consistent with previous research [2], we identified both information- and system or technology-related barriers, including a lack of support for interpreting medical information and issues with portal functionality, usability, and display of pertinent information. Patients with lower digital health literacy will continue to be underserved by digital transformation unless the barriers impeding access to, and understanding of, PHR information are addressed [40].

Interestingly, digital health literacy, as measured by the eHEALS score, did not predict patients' willingness to correct errors in their PHRs. This finding could be explained by the properties of the eHEALS scale, which was developed according to the definition of eHealth literacy 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" [45]. This definition was based on the first generation of static, read-only, health information technology (Web 1.0), which has led to the eHEALS being criticized as not being sufficiently comprehensive to measure the skills required to interact with and contribute to dynamic web-based health tools [46]. As suggested by one of the original authors of eHEALS, the content of the scale should be updated to reflect the modern complex, dynamic, and social nature of the web [47].

Policy Implications and Future Research

Establishing a feedback mechanism that encourages patients to identify and respond to PHR errors is aligned with the National Health Service Patient Safety Strategy, which sets out expectations for the involvement of patients, families, caregivers, and other lay people in providing safer care, allowing patients to move from being passive recipients of care to *vigilant stakeholders* [48,49]. However, patient-facing error surveillance systems would need to align with the diverse needs, preferences, and capabilities of patients, while also providing frontline staff and health care organizations with the opportunity to learn from errors [4,20,38]. Policy makers agree that more needs to be done to avoid marginalizing specific patient groups when implementing new digital health technologies [50]. An important first step is to ensure that the accessibility, functionality, and usability of patient portals meet the needs of all patients, including those who have difficulty interacting with digital content, nonnative English speakers, and other traditionally underserved groups. Ongoing development of PHR systems, incorporating channels for patient feedback on EHR accuracy, should integrate user-centered principles such that equal access and use can be achieved for all patients while also respecting individual choice and level of engagement [20,40]. Future research should seek to understand which categories of errors patients can reliably correct and examine the feasibility and safety of patient-facing error surveillance systems while continuing to address the impact of the digital divide and associated health inequalities.

Strengths and Limitations

A major strength of this study lies in the application of a mixed methods approach, which led to a comprehensive understanding of users' willingness and ability to be involved in PHR error surveillance in a diverse patient population. We ensured rigor and transparency in our qualitative analyses by applying the framework method to derive insights from patients' free-text responses [51]. We collected and analyzed a comprehensive set of patient characteristics, allowing us to explore classic demographic factors (age, sex, ethnicity, and educational level) in combination with important additional variables: health status, motivation to be involved in one's own care, and digital health literacy. This study included patients treated in North West London, which may limit the generalizability of our findings. However, of note, around one-third (145/445, 32.6%) of the survey respondents lived outside London, in other locations across England. Future studies should use geographical data to inform the relationship between social or material deprivation and patient portal engagement. This web-based survey study examined the views of patients who have access to and experience of using an electronic PHR, and the findings therefore reflect the views of a self-selected group of digitally empowered patients. Although we achieved a minimum sample size to ensure representativeness and adequate statistical power, the survey response rate was low. CIs for the association between patients' sociodemographic characteristics and the outcomes of interest were wide, indicating a potentially broader range of plausible predictors than could be detected in our study. Despite these limitations, our finding that digital health literacy predicts patients' understanding of portal information resonates with the current evidence and interest in reducing health inequalities.

Conclusions

Our findings demonstrate that patients are both able and willing to identify and respond to errors in their PHRs, although some barriers to understanding information in PHR systems persist and may disproportionately affect patients with lower digital health literacy. Further development of PHR systems, incorporating channels for patient feedback on the accuracy of their records, should integrate user-centered design principles such that equal access and engagement in mitigating errors in the record is possible for all patients.

Acknowledgments

This research was funded by the National Institute for Health Research (NIHR) Imperial Patient Safety Translational Research Centre and supported by the NIHR Imperial Biomedical Research Centre. This research was enabled by the Imperial Clinical Analytics Research and Evaluation environment and used the Imperial Clinical Analytics Research and Evaluation team and data resources. The views expressed are those of the authors and not necessarily those of the National Health Service, NIHR or Department of Health and Social Care.

Authors' Contributions

ALN, LF, MK, and EM designed the study. ALN and LF administered the survey. RL conducted the analyses, with contributions from LF for the qualitative analysis. RL drafted the manuscript. All authors have contributed to the revision, editing, and approval of the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Cross-tabulation of patients' sociodemographic characteristics and their understanding of the Care Information Exchange information.
[DOCX File , 28 KB - jmir_v24i7e37226_app1.docx]

Multimedia Appendix 2
Cross-tabulation of patients' sociodemographic characteristics and their preferences for responding to errors in their records.
[DOCX File , 28 KB - jmir_v24i7e37226_app2.docx]

Multimedia Appendix 3
The Strengthening the Reporting of Observational Studies in Epidemiology statement—checklist of items that should be addressed in reports of observational studies.
[DOCX File , 325 KB - jmir_v24i7e37226_app3.docx]

Multimedia Appendix 4
Sociodemographic characteristics of patients in the missing data and analysis samples.
[DOCX File , 25 KB - jmir_v24i7e37226_app4.docx]

Multimedia Appendix 5
Missing data analysis for structured questionnaire items.
[DOCX File , 24 KB - jmir_v24i7e37226_app5.docx]

References

1. Graber ML, Byrne C, Johnston D. The impact of electronic health records on diagnosis. Diagnosis (Berl) 2017 Nov 27;4(4):211-223. [doi: 10.1515/dx-2017-0012] [Medline: 29536944]
2. Freise L, Neves AL, Flott K, Harrison P, Kelly J, Darzi A, et al. Assessment of patients' ability to review electronic health record information to identify potential errors: cross-sectional web-based survey. JMIR Form Res 2021 Feb 26;5(2):e19074 [FREE Full text] [doi: 10.2196/19074] [Medline: 33635277]
3. Adams KT, Pruitt Z, Kazi S, Hettinger AZ, Howe JL, Fong A, et al. Identifying health information technology usability issues contributing to medication errors across medication process stages. J Patient Saf 2021 Dec 01;17(8):e988-e994. [doi: 10.1097/PTS.0000000000000868] [Medline: 34009868]
4. Bell SK, Delbanco T, Elmore JG, Fitzgerald PS, Fossa A, Harcourt K, et al. Frequency and types of patient-reported errors in electronic health record ambulatory care notes. JAMA Netw Open 2020 Jun 01;3(6):e205867 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.5867] [Medline: 32515797]
5. Weir CR, Hurdle JF, Felgar MA, Hoffman JM, Roth B, Nebeker JR. Direct text entry in electronic progress notes. An evaluation of input errors. Methods Inf Med 2003;42(1):61-67. [Medline: 12695797]
6. Bell SK, Gerard M, Fossa A, Delbanco T, Folcarelli PH, Sands KE, et al. A patient feedback reporting tool for OpenNotes: implications for patient-clinician safety and quality partnerships. BMJ Qual Saf 2017 Apr;26(4):312-322. [doi: 10.1136/bmjqs-2016-006020] [Medline: 27965416]
7. Herlihy M, Harcourt K, Fossa A, Folcarelli P, Golen T, Bell SK. An opportunity to engage obstetrics and gynecology patients through shared visit notes. Obstet Gynecol 2019 Jul;134(1):128-137. [doi: 10.1097/AOG.0000000000003309] [Medline: 31188333]
8. Bowman S. Impact of electronic health record systems on information integrity: quality and safety implications. Perspect Health Inf Manag 2013;10:1c [FREE Full text] [Medline: 24159271]
9. Cebul RD, Love TE, Jain AK, Hebert CJ. Electronic health records and quality of diabetes care. N Engl J Med 2011 Sep 01;365(9):825-833. [doi: 10.1056/NEJMsa1102519] [Medline: 21879900]
10. Classen DC, Holmgren AJ, Co Z, Newmark LP, Seger D, Danforth M, et al. National trends in the safety performance of electronic health record systems from 2009 to 2018. JAMA Netw Open 2020 May 01;3(5):e205547 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.5547] [Medline: 32469412]
11. Hussain MI, Reynolds TL, Zheng K. Medication safety alert fatigue may be reduced via interaction design and clinical role tailoring: a systematic review. J Am Med Inform Assoc 2019 Oct 01;26(10):1141-1149 [FREE Full text] [doi: 10.1093/jamia/ocz095] [Medline: 31206159]
12. Warren LR, Clarke J, Arora S, Darzi A. Improving data sharing between acute hospitals in England: an overview of health record system distribution and retrospective observational analysis of inter-hospital transitions of care. BMJ Open 2019 Dec 05;9(12):e031637 [FREE Full text] [doi: 10.1136/bmjopen-2019-031637] [Medline: 31806611]
13. Koppel R, Metlay JP, Cohen A, Abaluck B, Localio AR, Kimmel SE, et al. Role of computerized physician order entry systems in facilitating medication errors. JAMA 2005 Mar 09;293(10):1197-1203. [doi: 10.1001/jama.293.10.1197] [Medline: 15755942]
14. DesRoches CM, Bell SK, Dong Z, Elmore J, Fernandez L, Fitzgerald P, et al. Patients managing medications and reading their visit notes: a survey of opennotes participants. Ann Intern Med 2019 Jul 02;171(1):69-71. [doi: 10.7326/M18-3197] [Medline: 31132794]
15. Symons JD, Ashrafian H, Dunscombe R, Darzi A. From EHR to PHR: let's get the record straight. BMJ Open 2019 Sep 18;9(9):e029582 [FREE Full text] [doi: 10.1136/bmjopen-2019-029582] [Medline: 31537566]

16. Personal health record (PHR) - user insights. Royal College of Physicians. 2017 Jan 30. URL: https://www.rcplondon.ac.uk/projects/outputs/personal-health-record-phr-user-insights [accessed 2022-06-14]
17. Access to patient records through the NHS App. NHS Digital. URL: https://digital.nhs.uk/services/nhs-app/nhs-app-guidance-for-gp-practices/accelerating-patient-access-to-their-record [accessed 2022-01-13]
18. D'Costa SN, Kuhn IL, Fritz Z. A systematic review of patient access to medical records in the acute setting: practicalities, perspectives and ethical consequences. BMC Med Ethics 2020 Mar 02;21(1):18 [FREE Full text] [doi: 10.1186/s12910-020-0459-6] [Medline: 32122332]
19. Neves AL, Freise L, Laranjo L, Carter AW, Darzi A, Mayer E. Impact of providing patients access to electronic health records on quality and safety of care: a systematic review and meta-analysis. BMJ Qual Saf 2020 Dec;29(12):1019-1032 [FREE Full text] [doi: 10.1136/bmjqs-2019-010581] [Medline: 32532814]
20. Gerard M, Fossa A, Folcarelli PH, Walker J, Bell SK. What patients value about reading visit notes: a qualitative inquiry of patient experiences with their health information. J Med Internet Res 2017 Jul 14;19(7):e237 [FREE Full text] [doi: 10.2196/jmir.7212] [Medline: 28710055]
21. Mold F, de Lusignan S, Sheikh A, Majeed A, Wyatt JC, Quinn T, et al. Patients' online access to their electronic health records and linked online services: a systematic review in primary care. Br J Gen Pract 2015 Mar;65(632):e141-e151 [FREE Full text] [doi: 10.3399/bjgp15X683941] [Medline: 25733435]
22. Bell SK, Mejilla R, Anselmo M, Darer JD, Elmore JG, Leveille S, et al. When doctors share visit notes with patients: a study of patient and doctor perceptions of documentation errors, safety opportunities and the patient-doctor relationship. BMJ Qual Saf 2017 Apr;26(4):262-270 [FREE Full text] [doi: 10.1136/bmjqs-2015-004697] [Medline: 27193032]
23. Neves AL, Smalley KR, Freise L, Harrison P, Darzi A, Mayer EK. Determinants of use of the care information exchange portal: cross-sectional study. J Med Internet Res 2021 Nov 11;23(11):e23481 [FREE Full text] [doi: 10.2196/23481] [Medline: 34762063]
24. 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]
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. van der Vaart R, van Deursen AJ, Drossaert CH, Taal E, van Dijk JA, van de Laar MA. Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations. J Med Internet Res 2011 Nov 09;13(4):e86 [FREE Full text] [doi: 10.2196/jmir.1840] [Medline: 22071338]
27. Shiferaw KB, Mehari EA. Internet use and eHealth literacy among health-care professionals in a resource limited setting: a cross-sectional survey. Adv Med Educ Pract 2019;10:563-570 [FREE Full text] [doi: 10.2147/AMEP.S205414] [Medline: 31440113]
28. Milne RA, Puts MT, 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. [doi: 10.1007/s13187-014-0744-5] [Medline: 25355524]
29. Mitsutake S, Shibata A, Ishii K, Okazaki K, Oka K. [Developing Japanese version of the eHealth Literacy Scale (eHEALS)]. Nihon Koshu Eisei Zasshi 2011 May;58(5):361-371. [Medline: 21905612]
30. Hosmer D, Lemeshow S, Sturdivant R. Applied Logistic Regression. New York, NY, USA: John Wiley & Sons; 2013.
31. Zhang Z. Model building strategy for logistic regression: purposeful selection. Ann Transl Med 2016 Mar;4(6):111 [FREE Full text] [doi: 10.21037/atm.2016.02.15] [Medline: 27127764]
32. Snipes M, Taylor D. Model selection and Akaike information criteria: an example from wine ratings and prices. Wine Econ Policy 2014 Jun;3(1):3-9. [doi: 10.1016/j.wep.2014.03.001]
33. Ritchie J, Spencer L. Qualitative data for applied policy research. In: Bryman A, Burgess B, editors. Analyzing Qualitative Data. London, UK: Routledge; 1994:173-194.
34. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. PLoS Med 2007 Oct 16;4(10):e297 [FREE Full text] [doi: 10.1371/journal.pmed.0040297] [Medline: 17941715]
35. G7 patient access to health records: final report. Department of Health and Social Care, United Kingdom Government. 2021 Dec 30. URL: https://www.gov.uk/government/publications/g7-health-track-digital-health-final-reports/g7-patient-access-to-health-records-final-report [accessed 2022-06-14]
36. Online access to Digital GP records 2019/20. British Medical Association & NHS England. 2019. URL: https://www.england.nhs.uk/wp-content/uploads/2019/12/bma-nhse-joint-statement-prospective-record-access-v1.1.pdf [accessed 2022-05-25]
37. Ross SE, Lin C. The effects of promoting patient access to medical records: a review. J Am Med Inform Assoc 2003;10(2):129-138 [FREE Full text] [doi: 10.1197/jamia.m1147] [Medline: 12595402]
38. Lam BD, Bourgeois F, Dong ZJ, Bell SK. Speaking up about patient-perceived serious visit note errors: patient and family experiences and recommendations. J Am Med Inform Assoc 2021 Mar 18;28(4):685-694 [FREE Full text] [doi: 10.1093/jamia/ocaa293] [Medline: 33367831]

39. Johansen MA, Kummervold PE, Sørensen T, Zanaboni P. Health professionals' experience with patients accessing their electronic health records: results from an online survey. Stud Health Technol Inform 2019 Aug 21;264:504-508. [doi: 10.3233/SHTI190273] [Medline: 31437974]
40. Cheng C, Beauchamp A, Elsworth GR, Osborne RH. Applying the electronic health literacy lens: systematic review of electronic health interventions targeted at socially disadvantaged groups. J Med Internet Res 2020 Aug 13;22(8):e18476 [FREE Full text] [doi: 10.2196/18476] [Medline: 32788144]
41. Gerard M, Chimowitz H, Fossa A, Bourgeois F, Fernandez L, Bell SK. The importance of visit notes on patient portals for engaging less educated or nonwhite patients: survey study. J Med Internet Res 2018 May 24;20(5):e191 [FREE Full text] [doi: 10.2196/jmir.9196] [Medline: 29793900]
42. Sanders MR, Winters P, Fortuna RJ, Mendoza M, Berliant M, Clark L, et al. Internet access and patient portal readiness among patients in a group of inner-city safety-net practices. J Ambul Care Manage 2013;36(3):251-259. [doi: 10.1097/JAC.0b013e31829702f9] [Medline: 23748274]
43. Davis SE, Osborn CY, Kripalani S, Goggins KM, Jackson GP. Health literacy, education levels, and patient portal usage during hospitalizations. AMIA Annu Symp Proc 2015;2015:1871-1880 [FREE Full text] [Medline: 26958286]
44. 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]
45. 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]
46. van der Vaart R, Drossaert CH, de Heus M, Taal E, van de Laar MA. Measuring actual eHealth literacy among patients with rheumatic diseases: a qualitative analysis of problems encountered using Health 1.0 and Health 2.0 applications. J Med Internet Res 2013 Feb 11;15(2):e27 [FREE Full text] [doi: 10.2196/jmir.2428] [Medline: 23399720]
47. 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]
48. Implementing the NHS patient safety strategy: safer culture, safer systems, safer patients. NHS England & NHS Improvement. URL: https://www.england.nhs.uk/patient-safety/the-nhs-patient-safety-strategy/ [accessed 2022-05-25]
49. Framework for involving patients in patient safety. NHS England. 2021. URL: https://www.england.nhs.uk/patient-safety/framework-for-involving-patients-in-patient-safety/ [accessed 2022-05-25]
50. Digital clinical safety strategy. National Health Service. URL: https://www.nhsx.nhs.uk/key-tools-and-info/digital-clinical-safety-strategy/ [accessed 2022-05-25]
51. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013 Sep 18;13:117 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]

Abbreviations

CIE: Care Information Exchange
eHEALS: eHealth Literacy Scale
EHR: electronic health record
NIHR: National Institute for Health Research
OR: odds ratio
PHR: personal health record

Edited by T Leung; submitted 16.02.22; peer-reviewed by HL Tong, C Cheng; comments to author 02.05.22; revised version received 23.05.22; accepted 24.05.22; published 08.07.22.

Please cite as:
Lear R, Freise L, Kybert M, Darzi A, Neves AL, Mayer EK
Patients' Willingness and Ability to Identify and Respond to Errors in Their Personal Health Records: Mixed Methods Analysis of Cross-sectional Survey Data
J Med Internet Res 2022;24(7):e37226
URL: https://www.jmir.org/2022/7/e37226
doi:10.2196/37226
PMID:35802397

Original Paper

Inducing and Recording Acute Stress Responses on a Large Scale With the Digital Stress Test (DST): Development and Evaluation Study

Matthias Norden[1,2,3]; Amin Gerard Hofmann[1,2]; Martin Meier[2]; Felix Balzer[3]; Oliver T Wolf[4]; Erwin Böttinger[2,5]; Hanna Drimalla[1,2]

[1]Faculty of Technology, Bielefeld University, Bielefeld, Germany

[2]Digital Health Center, Hasso Plattner Institute, University of Potsdam, Potsdam, Germany

[3]Institute of Medical Informatics, Charité – Universitätsmedizin Berlin, Corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

[4]Department of Cognitive Psychology, Institute of Cognitive Neuroscience, Faculty of Psychology, Ruhr University Bochum, Bochum, Germany

[5]Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York City, NY, United States

Corresponding Author:
Hanna Drimalla
Faculty of Technology
Bielefeld University
Postfach 10 01 31
Bielefeld, 33619
Germany
Phone: 49 521 106 12043
Email: drimalla@techfak.uni-bielefeld.de

Abstract

Background: Valuable insights into the pathophysiology and consequences of acute psychosocial stress have been gained using standardized stress induction experiments. However, most protocols are limited to laboratory settings, are labor-intensive, and cannot be scaled to larger cohorts or transferred to daily life scenarios.

Objective: We aimed to provide a scalable digital tool that enables the standardized induction and recording of acute stress responses in outside-the-laboratory settings without any experimenter contact.

Methods: On the basis of well-described stress protocols, we developed the Digital Stress Test (DST) and evaluated its feasibility and stress induction potential in a large web-based study. A total of 284 participants completed either the DST (n=103; 52/103, 50.5% women; mean age 31.34, SD 9.48 years) or an adapted control version (n=181; 96/181, 53% women; mean age 31.51, SD 11.18 years) with their smartphones via a web application. We compared their affective responses using the international Positive and Negative Affect Schedule Short Form before and after stress induction. In addition, we assessed the participants' stress-related feelings indicated in visual analogue scales before, during, and after the procedure, and further analyzed the implemented stress-inducing elements. Finally, we compared the DST participants' stress reactivity with the results obtained in a classic stress test paradigm using data previously collected in 4 independent Trier Social Stress Test studies including 122 participants overall.

Results: Participants in the DST manifested significantly higher perceived stress indexes than the Control-DST participants at all measurements after the baseline (P<.001). Furthermore, the effect size of the increase in DST participants' negative affect (d=0.427) lay within the range of effect sizes for the increase in negative affect in the previously conducted Trier Social Stress Test experiments (0.281-1.015).

Conclusions: We present evidence that a digital stress paradigm administered by smartphone can be used for standardized stress induction and multimodal data collection on a large scale. Further development of the DST prototype and a subsequent validation study including physiological markers are outlined.

(J Med Internet Res 2022;24(7):e32280) doi:10.2196/32280

KEYWORDS

stress induction; smartphone; stress reactivity; Trier Social Stress Test; TSST; remote; video recording; acute stress; digital health; mobile health; mHealth; mobile phone

Introduction

Relevance and Rationale

Psychosocial stress is a major risk factor for the development of physical and mental illnesses, including hypertension, depression, and anxiety [1]. Valuable insights into its causes and consequences have been gained through experimental stress paradigms during which acute stressors are used to induce a psychosocial stress reaction. For example, such stress induction paradigms have been successfully used to investigate the effects of acute stress on the brain [2], hormonal and inflammatory reactivity [3], memory [4], and social cognition and behavior [5].

Applying controlled stress induction paradigms also enables the investigation of prevention and intervention strategies. For example, in a recent study, Het et al [6] used a classic stress paradigm to study the effects of an inpatient treatment on acute stress reactivity in women with eating disorders. In addition, controlled stress induction procedures play an important role in the development of objective stress detection methods [7,8] as they strongly rely on highly qualitative and representative data sets obtained through stress induction experiments [9].

Current Stress Paradigms and Their Limitations

Currently, most stress induction paradigms are limited in their scalability (ie, applicable across a large number of participants and distances) and, thus, cannot be easily used to gather large volumes of stress-related data. Furthermore, many of these have not been replicated outside the laboratory to verify the laboratory findings in outside-the-laboratory settings [10]. To overcome these limitations, a new standardized and validated stress induction paradigm is needed.

The Trier Social Stress Test (TSST) [11] is considered the gold standard in human experimental stress research, having been applied >4000 times including different populations and age groups [12]. Participants have to complete a 5-minute mock job interview and a 5-minute mental arithmetic task in front of an evaluating committee. This procedure requires a laboratory setup, an experimenter, and 2 actors playing the committee, making the TSST costly and unfeasible for large-scale application. In addition, the impact of the different methodological elements (eg, panel composition) on the stress reaction and the relatively small sample sizes complicate the reproducibility of the findings [13-15]. Furthermore, the experimental setting might lead to stress responses that differ from acute stress experienced in daily life.

Several adaptations have been made to provide less costly and laborious versions, but they still require human resources (eg, TSST for groups) or additional equipment (eg, virtual reality TSST or e-TSST) and have not been tested in nonlaboratory settings. Recently, 2 studies applied a web-based version of the TSST during which adolescent [16] or adult [17] participants joined judges and experimenters on a web-based videoconferencing platform without any in-person assessment. The responses to these web-delivered versions were consistent with standard in-person responses although the paradigm was conducted remotely. This highlights the possibility of assessing stress reactivity outside a research laboratory. However, the entire procedure still depends on live interactions between the committee, the participant, and the experimenter.

Stressors that enable the investigation of stress responses without direct experimenter contact have been developed for imaging scenarios [18]. Using their Imaging Paradigm for Evaluative Social Stress, Fehlner et al [19] showed that delivering short spoken answers to selected topics in front of a prerecorded audience and additional framings induced robust stress responses. This indicates that psychosocial stress can also be induced by making the participants believe they are exposed to some kind of social evaluation without direct experimenter interaction.

The Montreal Imaging Stress Task (MIST) [20] supports this assumption. It comprises computerized mental arithmetic tasks with an induced failure component and social pressure elements. However, these paradigms have only been tested within imaging laboratory settings where experimenters were still present and performed potentially stressful measurements. Thus, the stress induction might be influenced by the imaging setting and the experimenter's role during the procedure.

Many other well-described stress paradigms (eg, the CO2 challenge test and the socially evaluated cold pressor test [21]) are dependent on laboratory settings, build on physical stressors, and require human resources or additional equipment [22]. Other paradigms (eg, the Paced Auditory Serial Addition Task [23] and Stroop test [24]) would theoretically be applicable outside the laboratory but lack the possibility to collect multimodal behavior data (eg, facial expressions and voice recordings) of the stress response. To the best of our knowledge, there is currently no standardized and validated digital stress paradigm that can be carried out without an experimenter and collect multimodal video data of participants in stressed conditions. Therefore, we conceptualized and developed a completely digital stress test to address the need for an innovative, standardized, and validated stress induction protocol.

Digital Stress Test

The Digital Stress Test (DST) is primarily intended as a digital research tool. Importantly, we did not aim to develop a direct stress measurement or therapeutic tool. Instead, the DST enables researchers to gain additional insights into acute stress responses by making stress studies scalable and transferable to outside-the-laboratory settings and collecting stress-relevant video data at the same time. Thus, the DST is designed as an easy-to-use smartphone web application where participants conduct the study (via the internet) without any direct communication with researchers or additional resources required (ie, wearables or native app downloads).

It combines different well-known stress induction principles of classic stress paradigms and adapts them to a digital setting. According to a meta-analysis of psychological stress paradigms by Dickerson and Kemeny [25], a robust and reliable stress response can be induced by acute or chronic threats to social status, particularly when conditions are uncontrollable. Most likely, this would occur when failure or poor performance could reveal a lack of ability. Both principles have been proven effective in state-of-the-art stress paradigms and will be used as the basis for the digital stress induction paradigm.

Second, the DST aims to collect multimodal behavior data (ie, facial and voice cues) that can be used to build a basic data set for further (machine learning) analysis. Therefore, the embedded stress induction procedure will include a naturalistic speaking part (ie, comparable with daily speaking).

Objectives and Hypotheses

The aim of this study was to develop the first prototype of a DST web application and assess its feasibility as well as its stress induction potential. Therefore, we also provided a neutral version called the Control-DST (C-DST) that can be used similarly in web-based settings. We hypothesized that the DST would elicit a stronger stress response compared with the neutral condition. In addition, we placed our results in the context of previous studies conducting the gold-standard paradigm (TSST).

This paper is organized as follows: in the *Methods* section, we describe the concept and development of the DST and its control version. Furthermore, we provide details of the large-scale web-based study conducted to evaluate the feasibility and stress induction potential of the DST. In the *Results* section, we present statistical evidence for the stress induction potential of the DST. Finally, in the *Discussion* section, we discuss our results and potential limitations in light of previous work and outline plans for future research.

Methods

Concept and Development of the DST

We first describe the underlying stress induction paradigm as well as its adaptation for the development of a suitable control condition. We include illustrations of the first DST and C-DST prototypes and outline the technological aspects. Before starting the web-based evaluation study, we conducted a pilot study to finalize the prototypes based on participants' feedback.

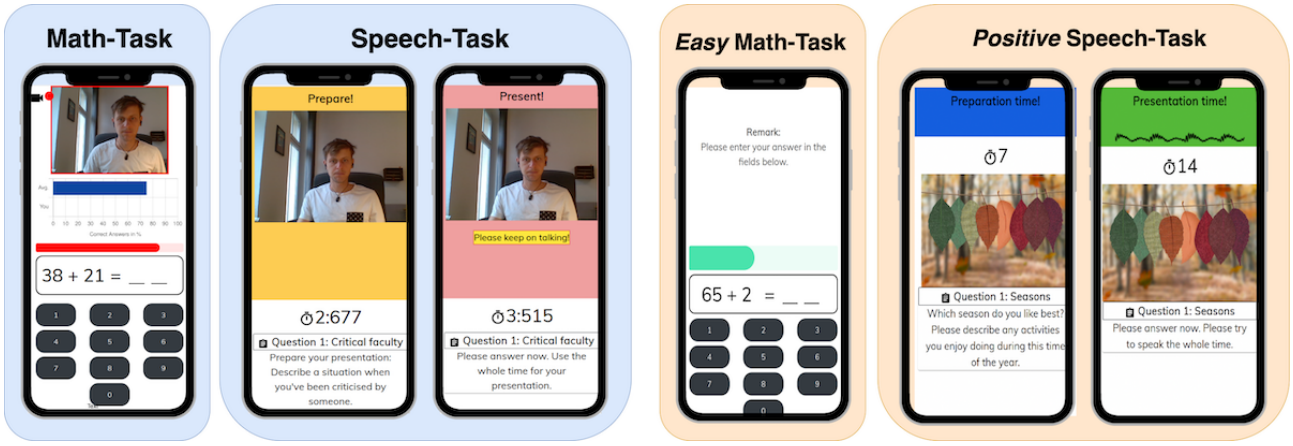
Concept of the DST

Overview

The paradigm consists of an arithmetic calculation and a free speech part and is framed as a cognitive-verbal performance test. Screenshots of the DST and its control condition are shown in Figure 1. The complete web application procedures can be seen in Multimedia Appendices 1 and 2. Presentation versions of the most recent DST and C-DST without any data saving can be found at their respective websites [26,27].

To elicit a robust acute psychosocial stress reaction, the DST procedure comprises multiple elements of social-evaluative threat and uncontrollability [11,25].

Figure 1. Screenshots of the Digital Stress Test (blue boxes) and Control - Digital Stress Test (orange boxes) tasks. The left box illustrates the Math-Task in each version, and the right box shows the preparation and presentation parts of the respective Speech-Task.



Framing

The DST is introduced as a research tool for "behavior analysis while performing a cognitive-verbal performance test," indicating that the individual performance of the participant is tested. To further increase the social-evaluative threat, they are informed that they will be recorded through the front camera of their smartphones and that these recordings are being analyzed to assess their individual resilience.

The participants record a short test video that claims to calibrate the implemented algorithm and shall increase the credibility of the automated behavior analysis.

The cognitive task is framed as a simple calculation task that a fake comparison group (based on age and gender) apparently solved with an average of 75% correct answers. This intends to emphasize the expected results and introduce the participants to the permanent social comparison in the upcoming calculation task, as done in the MIST [20].

Arithmetic Calculation Task (*Math-Task*)

The task comprises elements of the MIST [20] protocol adapted to the smartphone setting and enhanced with several other stress-inducing elements. After a countdown, the participants are required to solve simple calculation tasks consisting of addition, subtraction, multiplication, and division of 2 numbers

ranging from 1 to 99 with solutions ranging from 1 to 99. The participants need to type their solution on a number field within the given time limit. If the response is wrong or no response is recorded within the time limit, negative feedback is presented ("Wrong answer!" or "Too slow!") and the background color changes to red. After a correct response, the next calculation task is presented immediately. The time limit for each calculation is marked using a red expiring progress bar.

A continuous failure rate is being provoked. For the first task, the time limit is set to 3 seconds. If the participant answers a series of 3 consecutive arithmetic tasks correctly, the time limit is shortened by 10%. In addition, for the following 4 tasks, the numbers of the input field are swapped randomly to increase the difficulty and uncontrollability. If the participant answers a series of 3 consecutive tasks incorrectly (or not at all), the time limit is extended by 10%. If the participant does not give any input for 5 consecutive tasks, feedback indicating the relevance of the study is displayed, and the next task is chosen to be easily solvable (ie, a summation task). This intends to ensure ongoing participation.

During the Math-Task, the percentage of correct responses is continuously displayed and compared with the fixed average of the participants' age- and gender-related groups in a bar chart. As the achieved percentage of correct answers in the comparison group was claimed to be 75%, which usually exceeds the current percentage of the participant because of the implemented difficulty, this continually reminds the participant of failing.

In addition, the front camera of the smartphone is activated, and the recorded video is displayed directly on the upper half of the screen during the entire Math-Task. This intends to remind the participants that they are being recorded and apparently analyzed while failing in a school-like performance task.

The participants do not know how long the Math-Task takes to increase a feeling of uncontrollability. After 1.5 minutes, the Math-Task automatically stops. The participants see their final percentage of correct answers compared with the fabricated age- and gender-related average and are reminded that "only serious results can be used for this study," emphasizing the relevance of the participants' performance.

Free Speech Task (*Speech-Task*)

The second part is the Speech-Task, which further extends the social-evaluative threat through a presentation-like situation and enables the recording of stress-relevant voice cues. The participants are reminded that their verbal skills will be assessed. They are instructed to prepare structured and convincing verbal answers to standard job interview questions. They are not told how many questions will follow, making the length of this task unpredictable.

The Speech-Task includes 3 inconvenient answering scenarios (eg, "Describe a situation when you've been criticized by someone!") that are based on a previous study by Fehlner et al [19]. For each scenario, they are given 10 seconds to prepare and 20 seconds to present their speech. The participants are reminded to use the entire time for their presentation.

A countdown indicates the time for preparation and presentation, intending to pressure the participants. During their presentations, the background color of the entire screen blinks red to visually distract and agitate the participants.

The smartphone's front camera is activated, and the recorded video is displayed on the upper half of the screen during the preparation and presentation periods. In addition, a voice visualization is included in the presentation parts. After 1 second without recorded noise input signal, the participant is reminded to keep on talking, increasing the credibility of the behavior analysis and pressuring the participants, as done by the experimenters in the TSST paradigm [11,13]**.**

Concept of the C-DST

Overview

We also developed a control version of the DST that resembles its structure and procedure but differs in terms of the stress induction elements (Figure 1, right side). We changed the tasks and framings to be less stressful, as done for the placebo TSST [28] and friendly TSST [29]. The provided information on the study's background, privacy, and data protection aspects, as well as the performance task framing in the beginning, remains exactly the same to have a comparable baseline. The differences are outlined in the following sections.

Friendly Framing

The participants are informed that they are part of a control group and that they will not be video recorded. No recording of a test video or any additional framing of an automated behavior analysis takes place. The participants are not told that their individual performance results will be compared with those of other participants, and no fictive average result scores are displayed.

Easy Math-Task

The calculation tasks in the C-DST are generated in the same way as in the DST but only with summation tasks. The time limit for the first task is set to 5 seconds. The time adaptation algorithm is designed to enable more correct responses—as soon as the participant answers 1 task incorrectly (or not at all), the time limit is extended by 10%.

Only if the participant answers a series of 4 consecutive tasks correctly the time limit is shortened by 10%. In contrast to the DST, the numbers of the input field are not swapped for the following task.

The provided feedback is chosen to be encouraging (ie, the screen color changes to green for correct answers and does not change for wrong answers). The time limit is marked using a green progress bar. Neither a fake comparison with other participants' results nor any live recording through the front camera is displayed.

Positive Speech-Task

As opposed to the DST Speech-Task, the *positive* Speech-Task is not introduced as an assessment of the participants' verbal skills. The answering scenarios include only neutral topics (eg, "Which season do you like best? Please describe any activities you enjoy doing during this time of the year!").

Instead of displaying the live-recorded video of the front camera on the upper half of the screen, a neutral image suiting the question is shown. No further distraction through a red blinking background takes place, and the colors are chosen to be calming.

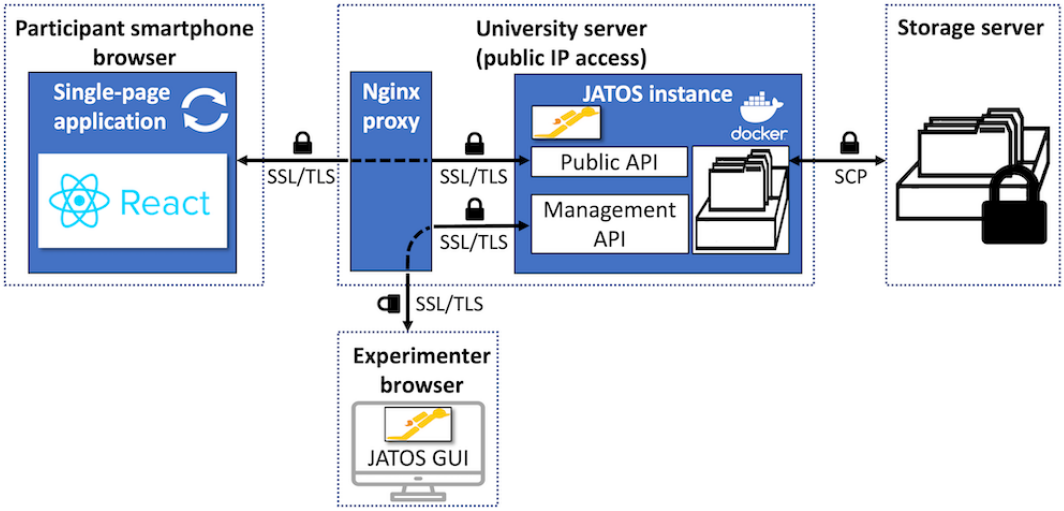
Technological Aspects

The system architecture of the applications is shown in Figure 2. The DST and C-DST were developed as single-page web applications using the JavaScript framework React.js. The source code of the most recent versions is publicly accessible at the website [30]. The applications run on standard browsers and are hosted on a university server that allows for public IP access using the open-source study management system JATOS [31] within a Docker container. JATOS exposes a public application programming interface (API) that is called with a wrapper library and handles requests from a participant's browser (eg, fetch and upload data). In addition, it provides a management API to handle requests from the experimenter's browser via the JATOS graphical user interface. More detailed information on the JATOS architecture can be found in the study by Lange et al [31].

In this study, only fully anonymized data were collected. We disabled the recording of videos but only streamed them within the participant's smartphone browser as the focus of this study was testing and validating the digital stress induction procedure. Owing to the capability and future plans to also collect sensitive and potentially identifiable video data, we implemented several security measures.

Nginx (Nginx, Inc) is used on the publicly reachable university server to ensure Secure Sockets Layer encryption, and it only responds to https requests for calls to both the public and management JATOS APIs. Participant data are only temporarily stored on the web server and directly transferred to a secure storage server via secure copy protocol after the test ends. All (remaining) data are deleted automatically from the web server in short time intervals. We have already received ethics approval for our data storage concept. For the future, we also plan to implement a client-side encryption of participant data files that takes place already within the web applications and can only be decrypted using private keys from the secure storage server.

Figure 2. System architecture of the test applications. The Digital Stress Test and Control - Digital Stress Test work as single-page web applications within the participant's smartphone browser (left side). The single-page applications are hosted on an web-based reachable university server (center) using the open-source web study management system JATOS [28] within a Docker container. API: application programming interface; GUI: graphical user interface; SCP: secure copy protocol; SSL: secure socket layer; TLS: transport layer security.



Pilot Study

We conducted a pilot study with 49 participants performing either the DST (21/49, 43%) or the C-DST (28/49, 57%) web application. On the basis of their feedback, we adjusted major usability issues that were caused by different browsers and smartphones and fixed technological bugs. We aimed for a comprehensive study introduction and consenting and debriefing information and modified the wording accordingly.

Evaluation of the DST

To assess the feasibility and stress induction potential of the DST, we first conducted a large web-based study. Participants in this web-based study performed either the DST or the C-DST and filled out several questionnaires regarding their affective responses. The effect sizes of the affective changes indicated by the DST participants in this web-based study were then compared with results obtained in previous studies performing the laboratory gold-standard paradigm (TSST).

Participants and Recruitment for the Web-Based Study

Overview

Participants were recruited via web-based publication of the study link in the university and study participation mailing lists, social networks (eg, Twitter and Facebook), podcasts, and websites. The study was conducted for 2 weeks, from February 10 to February 24, 2021. Within this period, 547 participants performed either the DST (300/547, 54.8%) or C-DST (247/547, 45.2%). For the evaluation of subjective stress parameters, we excluded participants with incomplete tests (229/547, 41.9%), previous self-reported knowledge of the framing (13/547, 2.4%), self-reported usability issues (19/547, 3.5%), or unrealistic procedure duration (2/547, 0.4%).

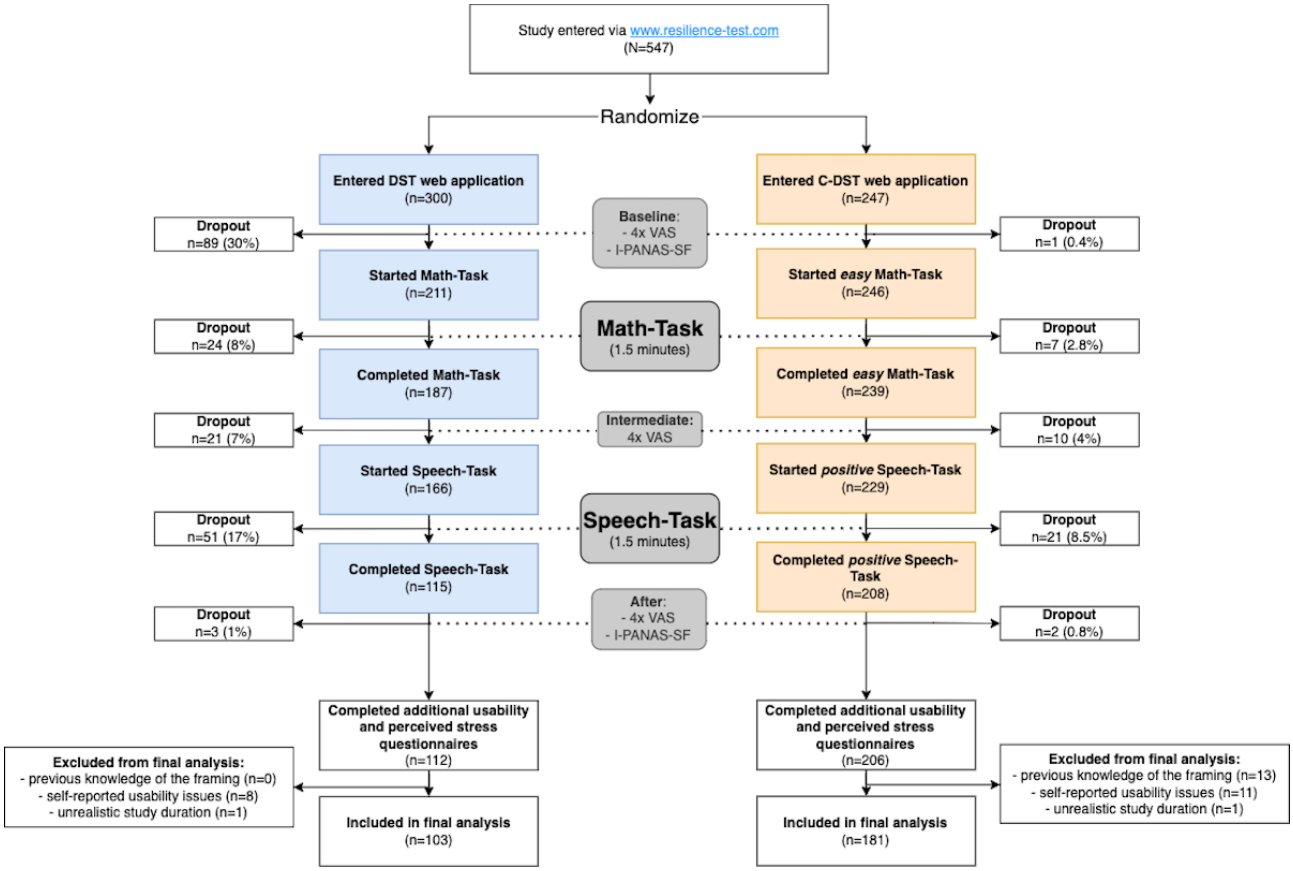
Web-Based Study Procedure

The design of the web-based study is shown in Figure 3. The entire procedure takes place on the screen of the participants' smartphones and takes approximately 5 to 10 minutes. Using the provided study link, the participants were randomly forwarded to either the stress or control paradigm. We adjusted the randomization algorithm to prefer the DST when we analyzed the dropout rates after the first week. However, most participants (479/547, 87.6%) performed the study within the first week because of the publication in a widespread German politics podcast.

During the pretest part of the web application, the participants were introduced to the study background, upcoming procedure, and privacy and data protection aspects. Assessments of the perceived stress level took place before, between, and after the 2 tasks using built-in questionnaires.

After completing both tasks, the participants were debriefed and linked to additional usability and follow-up questionnaires on an external website [32]. The participants could quit the study at any time (eg, by closing the browser).

Figure 3. Web-based study design, drop-outs, and collected data. The participants are randomly assigned to either the Digital Stress Test (blue boxes) or Control - Digital Stress Test (orange boxes) web application using the provided study link. The participants answer the same questionnaires within the respective web application and additional questionnaires on an external website afterward. I-PANAS-SF: international Positive and Negative Affect Schedule Short Form; VAS: Visual Analogue Scale.



Web-Based Study Data Collection

General demographic information, including age and gender, was obtained in the first part of the session. In addition, previous participation was asked about to exclude duplicated results with previous knowledge and confounding factors.

To assess the perceived stress levels, the participants completed several self-report questionnaires (Textbox 1).

To further compare the feasibility of the 2 paradigms, several pieces of metadata were stored during the procedure. This included the performance during the Math-Task and the study progress (ie, how much time the participants spent on different parts of the application and at which part they cancelled).

The participants were linked to additional questionnaires at the end of the web application. These included several questions on usability aspects (eg, problems with the correct camera and problems with the Visual Analogue Scale [VAS]) as well as the opportunity to provide open feedback. Furthermore, we asked the participants to rate their perceived stress regarding specific parts of the applications on a VAS.

Textbox 1. In-app self-report questionnaires completed by the participants.

Self-report questionnaires

- The international short form of the Positive and Negative Affect Schedule (PANAS) [33] was applied to assess positive and negative affect in the beginning (baseline) and after solving both tasks (posttest assessment). The PANAS is a well-validated and reliable tool to assess the participants' mood that has been applied in various studies on mood changes [34]. The participants indicate the intensity of 10 feelings and emotions on a 5-point Likert scale. The items can be subdivided into negative affect (NA; consisting of 5 items) and positive affect (PA; consisting of 5 items). We used the mean scores for both affects and normalized them for the number of items (ie, PA and NA outcomes ranging from 1 to 5 for each time point).
- Visual Analogue Scales (VAS) regarding 4 different dimensions of stress (feeling *stressed*, *frustrated*, *overstrained*, and *ashamed*) were obtained in the beginning, between the 2 tasks, and in the end. The VAS is a common instrument to measure characteristics that cannot be easily measured directly and is often used for pain, stress, or other subjective experiences [35]. The participants indicate how much they are perceiving specific feelings at the current moment by choosing a point on a fixed-size horizontal line where the ends are defined as the extremes (eg, *not at all* and *very much*). The VAS score is then determined by measuring the relative distance from the left end of the line to the participants' chosen point.

External Evaluation With the TSST

Data from 122 participants who underwent the traditional TSST procedure were previously collected at Ruhr University Bochum in 4 independent studies [34,36-38]. The procedures included assessments of the affective responses using the Positive and Negative Affect Schedule (PANAS). We used the archived data and compared the effects of the TSST on the participants' affective responses with the responses indicated by the DST participants in this study.

Statistical Analysis

Statistical analysis was performed using Python 3.7 (Python Software Foundation) with the pandas, statsmodels, and pingouin libraries. We assessed the distributions for normality and homogeneity of variances using Shapiro-Wilk and Levene tests, respectively. The participants' affective responses were analyzed using mixed-model ANOVAs for repeated measurements with the factor *time* (baseline and after for the PANAS; baseline, intermediate, and after for the VAS) and the between-subject factor *group* (DST group vs C-DST group) separately for the PANAS and VAS scales. Owing to their robustness against deviations from the normality assumption [39], we also used ANOVAs for nonnormally distributed data. Greenhouse-Geisser corrections for df were applied where sphericity could not be assumed. Post hoc tests were performed using Bonferroni-adjusted Welch *t* test for different sample sizes and nonhomogeneity of variances [40].

To further analyze the DST parts regarding their stress induction potential, we calculated the mean VAS scores for every part of the DST or C-DST evaluated in the posttest questionnaire and descriptively ranked them.

To compare the affective responses of the participants performing the DST with those of the participants who underwent the TSST in previous studies, we analyzed the normalized scores for the PANAS positive and negative affect subscales using a 2-step meta-analysis. Therefore, we first performed paired *t* tests on the normalized pre- and post-PANAS scores for each of the TSST studies separately and calculated standardized effect sizes. Afterward, we computed a combined effect size for all TSST studies by assigning weights based on the inverse of the change score variance to the individual effect sizes of the respective studies [41] and compared it with the standardized effect size observed in DST participants.

In all analyses reported, we used 2-tailed comparisons with a *P* value of <.05 as the significance criterion. The effect size was reported using partial η^2 for ANOVA and Cohen d_z for paired *t* tests [42].

Ethics Approval

Ethics approval for the study was granted by the University of Potsdam (application 33/2020), and the study was conducted in accordance with the General Data Protection Regulation. As this web-based study was conducted without experimenter supervision, special care was taken to ensure General Data Protection Regulation– and ethics-compliant informed consent, debriefing, and study cancellation process.

Results

Participants and Dropouts

Overall, 103 individuals completed the DST (50/103, 48.5% men; 52/103, 50.5% women; and 1/103, 1% other; mean age 31.34, SD 9.48 years), and 181 individuals completed the C-DST (83/181, 45.9% men, 96/181, 53% women, and 2/181, 1.1% other; mean age 31.51, SD 11.18 years). Most participants had a high level of education in both the DST (65/102, 63.7% had a university degree and 33/102, 32.4% had a high school degree) and C-DST groups (112/175, 64% had a university degree and 52/175, 29.7% had a high school degree). More details on the study participants' ages and educational backgrounds can be found in Figures S1 and S2 in Multimedia Appendix 3.

The average time taken to complete the procedure was 7.69 (SD 1.35) minutes for the DST and 6.53 (SD 1.05) minutes for the C-DST. Most participants (263/284, 92.6%) did not report any usability issues.

Beyond the completed studies, 247 individuals started the study but dropped out. For the C-DST, 83.4% (206/247) of the initial participants completed the procedure, whereas 37.3% (112/300) completed the DST paradigm. The dropout rates at different time points during the procedure are shown in Figure 3. Most DST participants who did not finish the study had already dropped out before starting the Math-Task. Participants who did not complete the study were not included in the following analyses.

DST Versus C-DST

The DST and C-DST participants' affective responses indicated in the PANAS questionnaires are shown in Figure 4. We found a significant main effect for the factor *group* ($F_{1,282}$=5.83; P=.02; η_p^2=0.02) accompanied by a significant *group*×*time* interaction effect ($F_{1,282}$=31.37; P<.001; η^2=0.10) in the PANAS negative affect subscale. Post hoc analyses for the *group* effect showed that the participants' overall reported negative affect was higher in the DST (mean 1.70, SD 0.55) than in the C-DST (mean 1.54, SD 0.57) group (P<.001). Post hoc tests for the *group*×*time* interaction effect revealed that the participants' indicated negative affect did not significantly differ between the DST (mean 1.57, SD 0.56) and C-DST (mean 1.58, SD 0.58) groups in the baseline measurements (P=.99) but was significantly higher in the posttest assessments for DST (mean 1.84, SD 0.7) than for C-DST (mean 1.49, SD 0.64) participants (P<.001).

Conducting separate mixed-model ANOVAs for the participants' indicated positive affect, we did not find significant *group* (P=.40) or *group*×*time* interaction (P=.51) effects, but we did find a significant *time* effect ($F_{1,282}$=0.43; P<.001; η_p^2=0.002). The post hoc analysis showed that, overall, perceived positive affect increased in study participants (baseline: 3.02 –0.65 to +0.65; after the procedure: 3.34 –0.76 to +0.76; P<.001).

The participants' responses to the 4 different VASs are shown in Figure 5. Regarding the *stress* scale, mixed-model ANOVAs revealed significant main effects for the factors *group* ($F_{1,282}$=14.42; P<.001; η_p^2=0.05) and *time* ($F_{2,564}$=75.11; P<.001; η_p^2=0.21) that were moderated by a significant *group*×*time* interaction effect ($F_{2,564}$=14.28; P<.001; η_p^2=0.05). Post hoc analyses for the *group* effect revealed that overall reported stress responses were higher for the DST (mean 42.39, SD 20.87) than for the C-DST (mean 32.79, SD 20.27) participants (P<.001). Post hoc tests for the *time* effect showed that participants' perceived stress significantly increased over the Math-Task (baseline: 32.92 –25.48 to +25.48; intermediate: 46.76 –26.01 to +26.01; P<.001) and decreased over the Speech-Task (intermediate: 46.76 –26.01 to +26.01; after the procedure: 29.15 –25.96 to +25.96; P<.001). Analyzing the *group*×*time* interaction, the participants' indicated VAS scores were significantly higher in the DST group than in the C-DST group at all time points after the baseline measurements (P<.001 in all cases). Furthermore, we found very similar patterns for the 3 other stress-related attributes (*frustration*, *shame*, and *overstrain*) conducting separate mixed-model ANOVAs and post hoc tests (Multimedia Appendix 4).

Figure 4. Negative (A) and positive (B) affect indicated in the Positive and Negative Affect Schedule (PANAS) subscales at baseline and posttest assessments for each participant in the Digital Stress Test (blue) and Control - Digital Stress Test (orange). A significant interaction between time and group was found for the negative but not the positive affect subscale. Digital Stress Test participants' negative affect was significantly higher at post-test assessment than Control - Digital Stress Test participants' negative affect (**P<.001 in post hoc Welch t test), whereas baseline scores did not significantly differ.

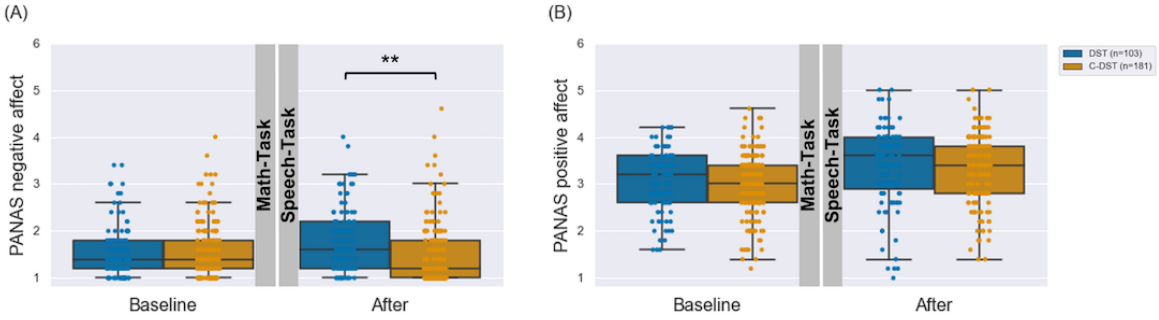
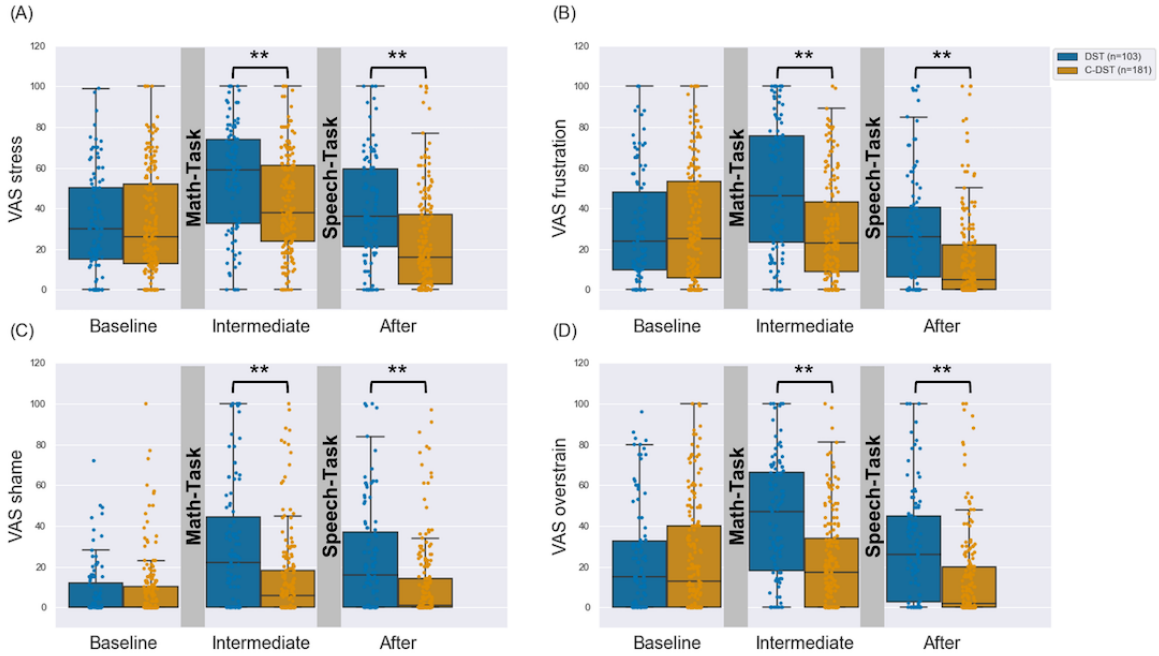


Figure 5. Visual Analogue Scale (VAS) responses for 4 different stress-related affect dimensions (A-D) of the Digital Stress Test (blue) and Control - Digital Stress Test (orange) groups at different times during the procedure. A significant interaction between time and group was found for all VAS scores. Subjective stress indexes were significantly elevated in the Digital Stress Test group compared with the Control Digital Stress test group at all time points after the baseline measurements (**P<.001 in post hoc Welch t test).



Analysis of Stress Elements

The results of the poststudy stress perception questionnaire are summarized in Table 1. According to the participants, the Math-Task was the most stressful element of the DST when compared with the framings in the beginning and with the Speech-Task.

Regarding the 2 tasks, the participants indicated the highest stress perception for the time pressure, whereas the social-evaluative component of being recorded through the front camera was not perceived as that stressful. Regarding the Math-Task, randomly swapping the input field after having correctly solved 3 calculation tasks seemed to induce a high level of perceived stress in the DST participants. Participants of the C-DST also rated the Math-Task and the implemented time limits as the most stressful elements of this version.

Table 1. Different parts of the Digital Stress Test (DST) and Control-Digital Stress Test (C-DST) and perceived stress levels sorted from highest to lowest indicating stress experience for each part of the DST paradigm.

Element in DST and C-DST and subcategory	Perceived stress—DST feedback (VAS[a]), mean (SD)	Perceived stress—C-DST feedback (VAS), mean (SD)
Framing		
Participation in a performance test	65.4 (23.8)	46.7 (25.2)
Behavior analysis through algorithm	40.6 (27.5)	N/A[b]
Math-Task (overall)	77.9 (18.4)	63.6 (23.7)
Time limit	88.8 (13.5)	73.3 (21.7)
Random input field swap	88.6 (15.6)	N/A
Feedback after every calculation task	67.3 (24.0)	40.8 (27.8)
Task difficulty	62.8 (26.3)	42.4 (23.0)
Live comparison with other participants	59.8 (28.5)	N/A
Personal performance	58.0 (29.0)	29.4 (22.3)
Front camera activation	39.7 (28.6)	N/A
Speech-Task (overall)	46.2 (23.3)	25.5 (19.6)
Preparation periods	45.5 (25.8)	19.6 (22.4)
Time limits	45.1 (27.3)	22.6 (24.8)
Questions	43.6 (23.8)	20.2 (22.7)
Front camera activation	37.1 (25.9)	N/A
Audio visualization of the voice	30.2 (24.2)	18.0 (21.8)

[a]VAS: Visual Analogue Scale.

[b]N/A: not applicable.

DST Versus TSST

To further evaluate the stress induction potential of the DST, we performed a 2-step meta-analysis and compared the effects of the DST with findings of 4 previously conducted TSST studies. The results of each study are shown in Table 2. The sample sizes of the TSST studies ranged from 20 to 50, whereas 103 participants completed the DST in this study.

The participants' indicated negative affect significantly increased in the DST and in all but one of the TSST studies (Table 2). The standardized effect sizes for the change in negative affect in the TSST studies ranged from 0.281 to 1.015, with a combined effect size of 0.667. The calculated effect for the increase in negative affect in the DST participants was 0.427.

The reported positive affect significantly increased in the DST participants in this study, whereas the results of the 4 TSST studies did not reveal significant changes in positive affect (Table 2). The standardized effect sizes for the change in positive affect in the TSST studies ranged from 0.022 to 0.363, with a combined effect size of 0.119. The calculated effect for the increase in positive affect in the DST participants was 0.382.

Table 2. Overview of studies used for meta-analytical comparison of the Digital Stress Test (DST) with the Trier Social Stress Test (TSST) effect including paired *t* test results for each study.

Study and PANAS[a] subscale	Baseline score, mean (SD)	Score after, mean (SD)	Change, mean (SD)	Paired *t* test		
				t test (*df*)	*P* value	Cohen d_z
DST (n=103)						
NA[b]	1.57 (0.56)	1.84 (0.70)	0.27 (0.61)	−4.51 (102)	*<.001* [c]	0.427
PA[d]	3.08 (0.65)	3.37 (0.85)	0.29 (0.61)	−4.84 (102)	*<.001*	0.382
TSST [37] (n=26)						
NA	1.28 (0.38)	1.82 (0.66)	0.55 (0.68)	−4.08 (25)	*<.001*	1.015
PA	2.95 (0.45)	2.94 (0.59)	−0.01 (0.51)	0.11 (25)	.91	0.022
TSST [34] (n=26)						
NA	1.33 (0.32)	1.66 (0.57)	0.32 (0.44)	−3.75 (25)	*<.001*	0.694
PA	2.95 (0.56)	2.97 (0.64)	0.02 (0.45)	−0.17 (25)	.86	0.026
TSST [36] (n=20)						
NA	1.36 (0.33)	1.51 (0.71)	0.16 (0.68)	−1.02 (19)	.32	0.281
PA	2.75 (0.42)	3.02 (0.96)	0.27 (0.82)	−1.47 (19)	.16	0.363
TSST [38] (n=50)						
NA	1.43 (0.56)	1.85 (0.72)	0.42 (0.53)	−5.64 (49)	*<.001*	0.655
PA	3.02 (0.57)	2.88 (0.68)	−0.14 (0.47)	2.10 (49)	.04	0.221

[a]PANAS: Positive and Negative Affect Schedule.

[b]NA: negative affect.

[c]Italics emphasize significance.

[d]PA: positive affect.

Discussion

Principal Findings

In this proof-of-concept study, we evaluated the feasibility of a fully digitalized acute stress paradigm for smartphones, the DST, to induce and record psychosocial stress responses in outside-the-laboratory settings. We compared it with a digital control condition (C-DST) in a large web-based study and set the effect size of the participants' indicated affect changes in the context of results previously achieved in the TSST. To our knowledge, this is the first study evaluating the stress reactivity of an experimenter-independent paradigm that does not include any human-human interaction.

We showed that the DST significantly induced higher levels of perceived stress and negative affect than the control condition. In addition to feeling more stressed, DST participants also reported similar increases in related affects such as frustration, shame, and overstrain. Notably, the reported increases in negative affect indicated by DST participants not only significantly exceeded those of participants performing the C-DST but were also comparable with those reported by TSST participants in previous studies regarding the calculated effect sizes.

These findings provide convincing evidence that an acute psychosocial stress response can be induced with a smartphone without any further equipment or experimenters taking part. In particular, the DST managed to induce subjective stress even if the social-evaluative threat and uncontrollability [25] of this study can be assumed to be weaker than in previous studies. TSST participants performed the paradigm in the laboratory and were administered physiological measurements and watched by several experimenters, whereas the DST and C-DST were mainly performed at home without any additional procedures or people present. Participants in the web-based study took <10 minutes for the whole paradigm and could cancel the study at any time by simply closing the browser. Nevertheless, the mere framing of social evaluation, a difficult mathematical task, and a free speech task in front of the smartphone camera were sufficient to elicit a psychological stress response.

These findings extend the results obtained in other studies analyzing the stress induction potential of less controlled and experimenter-dependent stress paradigms. Virtual reality versions of the TSST successfully elicit psychosocial stress responses using prerecorded [43], animated [44-47], or even nonhuman robot audiences [48]. However, these protocols still require experimenters to conduct the procedure.

Although previous studies have focused on the development of more immersive and convincing virtual realities to improve stress induction [49], our results indicate that the procedure might be simplified and spare human-human interaction. The recently investigated internet-delivered TSST has already shown that a significant stress response can be induced without direct

person-to-person contact [16,17]. Our study supports these findings and further leads to the assumption that psychosocial stress can be induced without any live interaction.

Interestingly, in addition to evoking a significant level of perceived stress and negative affect, the DST also increased the participants' positive affect. Increases in positive affect have also been reported in other studies, including stress tests [36,37]. We assume that the increase in this study was caused by an end-of-study relief and self-selection bias. First, the participants in this web-based study knew that the performance test would end after the last questionnaire, whereas, in many other studies, experimental measures or interventions followed the stress paradigm [14,50,51]. Second, participants with a strong decrease in their positive affect might have cancelled the study because of the very low cancellation barrier.

For a more detailed investigation of the stress induction potential of our new paradigm, we also examined the elements implemented for stress induction in the DST regarding the participants' responses. Previous work has highlighted the impact of social evaluation and unpredictability on stress response. In particular, public speaking parts have been shown to induce stress in participants [25,52,53]. In our study, we found a strong increase in perceived stress throughout the Math-Task and a subsequent slight decrease over the Speech-Task. The results of the posttest questionnaires also indicate that the participants perceived the Math-Task as the more stressful task. In contrast to the TSST, the Speech-Task was the last part of the procedure, and the participants knew that the study would end afterward. Thus, the affect ratings might also have been influenced by the task order and end-of-study relief. Another reason for the lower stress induced by the Speech-Task might be that speaking to the front camera without any real social evaluation does not induce as much stress as that experienced in live experiments. Similarly, other paradigms that include a social-evaluative stressor without direct human interaction resulted in weaker stress responses [44,46,48]. In addition, despite receiving live feedback from the audio input, the participants might not talk or might skip the task as there is no real experimenter control. Furthermore, the participants in this study knew that their recordings would not be saved or watched.

For upcoming web-based studies, permanently saving the videos and the possibility that experimenters watch them might increase psychosocial stress. In addition, improving the credibility of the automated analysis through the implementation of more sophisticated adaptive feedback might lead to a stronger feeling of social evaluation. Another approach to strengthen the social-evaluative characteristics of the DST could be to implement a prerecorded or animated audience instead of displaying the participants' own video recordings. In addition, strengthening the social comparison characteristics of the paradigm through fabricated comparisons of the performance during the Speech-Task (similar to the Math-Task) might lead to a stronger psychosocial stress induction.

Web-Based Feasibility of the DST

The evaluation of the DST in a web-based study highlights the potential of this paradigm. Within 2 weeks, nearly 600 participants performed one of the versions, and almost 300 completed it. By contrast, a recent review evaluating 35 TSST studies showed that the average number of participants was 47, with only 1 study including >100 participants [54]. Campbell and Ehlert [55] evaluated 359 TSST and TSST-related articles and found only 6 studies that reported >100 participants, presuming many more laborious and time-consuming studies. Even in the recently proposed web-based TSST, experimenters and actors need to be present during the web-based videoconferencing session, and the still laborious procedure is stated as a limitation by the authors [16].

Another advantage of the DST procedure is its inclusiveness, allowing for participation from any location and in different conditions. However, the number and composition of the participants highly depend on the recruitment process. Many participants entered this study because of its announcement in a well-known German political podcast and university mailing lists, which might have led to age and educational background selection bias in our sample. The participants in this study were mainly younger and from higher educational backgrounds. Previous studies have shown differences in stress reactivity according to age and socioeconomic status, which need to be addressed when interpreting the findings of this study. In some studies, physiological stress responses to cognitive challenges were stronger in older and higher-educated individuals [56-58]. Nicolson et al [59] found stronger cortisol reactivity in younger individuals and no age-related differences in emotional responses to a speech task. According to Dickerson et al [25], cognitive testing may be more stressful for older adults with higher levels of education as they perceive a greater threat of negative social evaluation. Moreover, the average lower digital literacy of older adults [60] may even increase the stress response in older participants in a smartphone-based paradigm such as the DST. However, future studies should verify the stress induction potential for individuals of other ages and educational backgrounds.

In a web-based study without any direct supervision, it is crucial to ensure that the participants follow the correct procedure of the experiment. Therefore, the participants were automatically reminded to continue when they did not react during the tasks for a certain time. In addition, we logged the study progress and excluded participants who were extreme outliers with respect to the study duration. In the future, we plan to also analyze the video recordings regarding compliance and include more detailed live feedback.

The barrier for dropping out of this web-based study was much lower than that in laboratory or other live-contact settings. The participants could cancel the study at any time simply by closing the browser of their smartphones. Although it was ethically favorable that participants did not need to continue when they felt overwhelmed by the test situation, this also affected the outcome of the study. Many participants (324/547, 59.2%) dropped out even though the procedure took <10 minutes and no personal data were saved permanently. Most DST participants (89/300, 29.7%) had already cancelled during the introduction, which was not observed in the C-DST group. We assume that the higher cancellation rate in the beginning was caused by

technological problems or privacy concerns related to the video recording in the DST.

For future versions, we plan to emphasize the high standard of data protection implemented in the DST and a cancellation procedure that allows for further decision-making regarding the submitted data and short feedback on the cancellation reasons.

Limitations

Previous studies have highlighted the long-term consequences of acute stress–induced physiological changes [61], which were not evaluated in this study. Although, in some experiments, correlations between psychological and physiological stress responses could be found [62,63], others could not verify this [25,55,64]. Hellhammer and Schubert [65] found that psychological measurements during, but not before or after, the TSST were related to physiological responses. The DST participants reported the highest level of perceived stress between the 2 tasks, indicating that physiological changes might also have taken place. Even if it is not yet clear whether the stress response elicited by the DST entails physiological changes, addressing psychological stress reactivity plays an important role in the individual quality of life [66] and mental well-being [67]. Previous studies have shown the effects of interventions on psychological well-being [68,69], which highlights the potential use of the DST for evaluating stress intervention strategies.

Nevertheless, the stress induction potential of the DST should be confirmed in a follow-up study including measurements of other stress-relevant systems such as the sympathetic nervous system and the hypothalamic-pituitary-adrenal axis [55,70].

Future Research

Several improvements to the stress induction procedure as well as the usability have been outlined. In particular, additional adaptive feedback algorithms that react to the participants' live-recorded behavior might improve the credibility of the social-evaluative framing and enhance compliance.

To further validate the DST, we plan to compare the psychological and physiological stress responses, including cortisol, heart rate, and blood pressure measurements, of participants undergoing the TSST and DST in a within-subject design. Next, to improve and validate the stress induction procedure, we aim to adjust and evaluate the video data collection in the DST and build a large data set of stress test videos.

The DST might then be easily applied to different (clinical) cohorts (eg, stress in patients with chronic pain [71], stress in patients with cancer [72], and stress in students [73]) and contexts (eg, job stress [74] and parental stress [75]) from any internet-connected location worldwide. In contrast to existing protocols, this would also allow for the conduction of stress studies in outside-the-laboratory scenarios and with individuals from diverse cultural, ethnic, and geographical backgrounds (eg, remote cultures) [76].

In contrast, the multimodal video data collected using the DST could serve as the basis for the development of video-based stress analysis algorithms using machine learning methods [77]. Baird et al [78] combined 3 data sets including videos and voice recordings of participants undergoing the TSST in separate studies for the prediction of acute stress responses. Consequently, the data obtained with the DST could enrich existing video data sets and be used in combination with them (eg, pretraining for personalized models [79] and cross-model transfer learning [78]) to improve the quality of the algorithms. From a more long-term perspective, these algorithms might be used within the DST to provide feedback on a participant's stress reactivity and evaluate personal prevention or intervention strategies (eg, resilience trainings [80]).

Conclusions

To the best of our knowledge, this is the first approach to a standardized digital stress paradigm that can be carried out using only a smartphone. Moreover, our results imply that psychosocial stress can be induced through cognitive-verbal performance tasks and additional framings in a fully automated web application.

The ability to conduct (stress) studies without any experimenter or additional equipment required can also be seen as a potential turning point for translating traditional (stress) research to the wild. Owing to the web application–based mobile architecture, future researchers can quickly prepare, conduct, adapt, and evaluate studies anywhere—including basic and clinical research. In accordance with the principles of open access, the source code of the DST and C-DST is publicly available, and both applications can be freely used for research purposes upon request.

Future studies will evaluate the potential of the implemented video recording capability to provide a high-quality stress data set for algorithm development. This study may serve as inspiration to bridge the gap between classic psychological research and interdisciplinary computer science.

Acknowledgments

This study was funded by the Federal Ministry of Education and Research (award 01IS20046). The Federal Ministry of Education and Research had no involvement in the study design; the collection, analysis, and interpretation of the data; the writing of the report; or the decision to submit the paper for publication.

Conflicts of Interest

FB reports grants from the German Federal Ministry of Education and Research, grants from the German Federal Ministry of Health, grants from the Berlin Institute of Health, personal fees from Elsevier Publishing and Medtronic, grants from the Hans

Böckler Foundation, other grants from the Robert Koch Institute, grants from the Einstein Foundation, and grants from Berlin University Alliance outside the submitted work.

Multimedia Appendix 1
Video illustrating the procedure for the Digital Stress Test web application used in this study. The video has been modified to 2x playback speed, and recorded sound has been removed for publication purposes. For proper inspection or reading of specific parts, the video can be paused. Screen recordings were done by a research assistant in May 2021. The mock-up design was provided by Vectorium - de.freepik.com. A presentation version of the most recent digital stress test version without permanent data saving can be found at www.digitalstresstest.org.
[MP4 File (MP4 Video), 60338 KB - jmir_v24i7e32280_app1.mp4]

Multimedia Appendix 2
Video illustrating the procedure for the Control - Digital Stress Test web application used in this study. The video has been modified to 2x playback speed, and recorded sound has been removed for publication purposes. For proper inspection or reading of specific parts, the video can be paused. Screen recordings were done by a research assistant in May 2021. The mock-up design was provided by Vectorium - de.freepik.com. A presentation version of the most recent control digital stress test version without permanent data saving can be found at www.digitalstresstest.org/control.
[MP4 File (MP4 Video), 44183 KB - jmir_v24i7e32280_app2.mp4]

Multimedia Appendix 3
Distribution of age (Figure S1) and educational background (Figure S2) across Digital Stress Test and Control - Digital Stress Test participants.
[PDF File (Adobe PDF File), 72 KB - jmir_v24i7e32280_app3.pdf]

Multimedia Appendix 4
Mixed-model ANOVA results for the comparison of participants' affective responses indicated in the different VAS and PANAS questionnaires at different time points. Second sheet displays post-hoc analyses for significant effects.
[XLS File (Microsoft Excel File), 20 KB - jmir_v24i7e32280_app4.xls]

References

1. Cohen S, Janicki-Deverts D, Miller GE. Psychological stress and disease. JAMA 2007 Oct 10;298(14):1685-1687. [doi: 10.1001/jama.298.14.1685] [Medline: 17925521]
2. Dedovic K, D'Aguiar C, Pruessner JC. What stress does to your brain: a review of neuroimaging studies. Can J Psychiatry 2009 Jan;54(1):6-15. [doi: 10.1177/070674370905400104] [Medline: 19175975]
3. Chen X, Gianferante D, Hanlin L, Fiksdal A, Breines JG, Thoma MV, et al. HPA-axis and inflammatory reactivity to acute stress is related with basal HPA-axis activity. Psychoneuroendocrinology 2017 Apr;78:168-176 [FREE Full text] [doi: 10.1016/j.psyneuen.2017.01.035] [Medline: 28209543]
4. Wolf OT. Stress and memory retrieval: mechanisms and consequences. Curr Opin Behav Sci 2017 Apr;14:40-46. [doi: 10.1016/j.cobeha.2016.12.001]
5. von Dawans B, Strojny J, Domes G. The effects of acute stress and stress hormones on social cognition and behavior: current state of research and future directions. Neurosci Biobehav Rev 2021 Feb;121:75-88. [doi: 10.1016/j.neubiorev.2020.11.026] [Medline: 33301780]
6. Het S, Vocks S, Wolf JM, Herpertz S, Wolf OT. Treatment-resistant blunted HPA activity, but reversible cardiovascular stress reactivity in young women with eating disorders. Front Psychiatry 2020 Jul 22;11:726 [FREE Full text] [doi: 10.3389/fpsyt.2020.00726] [Medline: 32793011]
7. Can YS, Arnrich B, Ersoy C. Stress detection in daily life scenarios using smart phones and wearable sensors: a survey. J Biomed Inform 2019 Apr;92:103139 [FREE Full text] [doi: 10.1016/j.jbi.2019.103139] [Medline: 30825538]
8. Pisanski K, Kobylarek A, Jakubowska L, Nowak J, Walter A, Błaszczyński K, et al. Multimodal stress detection: testing for covariation in vocal, hormonal and physiological responses to Trier Social Stress Test. Horm Behav 2018 Nov;106:52-61. [doi: 10.1016/j.yhbeh.2018.08.014] [Medline: 30189213]
9. Mahesh B, Hassan T, Prassler E, Garbas JU. Requirements for a reference dataset for multimodal human stress detection. In: Proceedings of the 2019 IEEE International Conference on Pervasive Computing and Communications Workshops. 2019 Presented at: PerComW '19; March 11-15, 2019; Kyoto, Japan p. 492-498. [doi: 10.1109/percomw.2019.8730884]
10. Picard RW. Automating the recognition of stress and emotion: from lab to real-world impact. IEEE MultiMedia 2016 Aug 5;23(3):3-7. [doi: 10.1109/mmul.2016.38]
11. Kirschbaum C, Pirke KM, Hellhammer DH. The 'Trier Social Stress Test'—a tool for investigating psychobiological stress responses in a laboratory setting. Neuropsychobiology 1993;28(1-2):76-81. [doi: 10.1159/000119004] [Medline: 8255414]

12. Allen AP, Kennedy PJ, Dockray S, Cryan JF, Dinan TG, Clarke G. The Trier Social Stress Test: principles and practice. Neurobiol Stress 2017 Feb;6:113-126 [FREE Full text] [doi: 10.1016/j.ynstr.2016.11.001] [Medline: 28229114]
13. Birkett MA. The Trier Social Stress Test protocol for inducing psychological stress. J Vis Exp 2011 Oct 19(56):3238 [FREE Full text] [doi: 10.3791/3238] [Medline: 22042290]
14. Goodman WK, Janson J, Wolf JM. Meta-analytical assessment of the effects of protocol variations on cortisol responses to the Trier Social Stress Test. Psychoneuroendocrinology 2017 Jun;80:26-35. [doi: 10.1016/j.psyneuen.2017.02.030] [Medline: 28292684]
15. Kudielka BM, Hellhammer DH, Kirschbaum C. Ten years of research with the trier social stress test—revisited. In: Harmon-Jones E, Winkielman P, editors. Social Neuroscience: Integrating Biological and Psychological Explanations of Social Behavior. New York, NY, USA: The Guilford Press; 2007:56-83.
16. Gunnar MR, Reid BM, Donzella B, Miller ZR, Gardow S, Tsakonas NC, et al. Validation of an online version of the Trier Social Stress Test in a study of adolescents. Psychoneuroendocrinology 2021 Mar;125:105111 [FREE Full text] [doi: 10.1016/j.psyneuen.2020.105111] [Medline: 33341502]
17. Eagle DE, Rash JA, Tice L, Proeschold-Bell RJ. Evaluation of a remote, Internet-delivered version of the Trier Social Stress Test. Int J Psychophysiol 2021 Jul;165:137-144. [doi: 10.1016/j.ijpsycho.2021.03.009] [Medline: 33865901]
18. Berretz G, Packheiser J, Kumsta R, Wolf OT, Ocklenburg S. The brain under stress-a systematic review and activation likelihood estimation meta-analysis of changes in BOLD signal associated with acute stress exposure. Neurosci Biobehav Rev 2021 May;124:89-99. [doi: 10.1016/j.neubiorev.2021.01.001] [Medline: 33497786]
19. Fehlner P, Bilek E, Harneit A, Böhringer A, Moessnang C, Meyer-Lindenberg A, et al. Neural responses to social evaluative threat in the absence of negative investigator feedback and provoked performance failures. Hum Brain Mapp 2020 Jun 01;41(8):2092-2103 [FREE Full text] [doi: 10.1002/hbm.24932] [Medline: 31958212]
20. Dedovic K, Renwick R, Mahani NK, Engert V, Lupien SJ, Pruessner JC. The Montreal Imaging Stress Task: using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain. J Psychiatry Neurosci 2005 Sep;30(5):319-325 [FREE Full text] [Medline: 16151536]
21. Schwabe L, Haddad L, Schachinger H. HPA axis activation by a socially evaluated cold-pressor test. Psychoneuroendocrinology 2008 Jul;33(6):890-895. [doi: 10.1016/j.psyneuen.2008.03.001] [Medline: 18403130]
22. Bali A, Jaggi AS. Clinical experimental stress studies: methods and assessment. Rev Neurosci 2015;26(5):555-579. [doi: 10.1515/revneuro-2015-0004] [Medline: 26020552]
23. Lejuez CW, Kahler CW, Brown RA. A modified computer version of the Paced Auditory Serial Addition Task (PASAT) as a laboratory-based stressor. Behav Ther 2003;26(4):290-293.
24. Renaud P, Blondin JP. The stress of Stroop performance: physiological and emotional responses to color–word interference, task pacing, and pacing speed. Int J Psychophysiol 1997 Sep;27(2):87-97. [doi: 10.1016/s0167-8760(97)00049-4]
25. Dickerson SS, Kemeny ME. 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]
26. Digital Stress Test. URL: https://www.digitalstresstest.org [accessed 2022-06-29]
27. Digital Stress Test Control. URL: https://www.digitalstresstest.org/control [accessed 2022-06-29]
28. Het S, Rohleder N, Schoofs D, Kirschbaum C, Wolf OT. Neuroendocrine and psychometric evaluation of a placebo version of the 'Trier Social Stress Test'. Psychoneuroendocrinology 2009 Aug;34(7):1075-1086. [doi: 10.1016/j.psyneuen.2009.02.008] [Medline: 19307062]
29. Wiemers US, Schoofs D, Wolf OT. A friendly version of the trier social stress test does not activate the HPA axis in healthy men and women. Stress 2013 Mar;16(2):254-260. [doi: 10.3109/10253890.2012.714427] [Medline: 22813431]
30. Digital Stress Test Code. URL: https://www.digitalstresstest.org/code [accessed 2022-06-29]
31. Lange K, Kühn S, Filevich E. "Just Another Tool for Online Studies" (JATOS): an easy solution for setup and management of Web servers supporting online studies. PLoS One 2015 Jun 26;10(6):e0130834 [FREE Full text] [doi: 10.1371/journal.pone.0130834] [Medline: 26114751]
32. Leiner DJ. SoSci Survey (Version 3.2.33). SoSci Survey. 2014. URL: https://www.soscisurvey.de [accessed 2021-02-01]
33. Thompson ER. Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS). J Cross Cult Psychol 2007 Mar 1;38(2):227-242. [doi: 10.1177/0022022106297301]
34. Herten N, Otto T, Wolf OT. The role of eye fixation in memory enhancement under stress - an eye tracking study. Neurobiol Learn Mem 2017 Apr;140:134-144. [doi: 10.1016/j.nlm.2017.02.016] [Medline: 28267591]
35. Lesage FX, Berjot S, Deschamps F. Clinical stress assessment using a visual analogue scale. Occup Med (Lond) 2012 Dec;62(8):600-605. [doi: 10.1093/occmed/kqs140] [Medline: 22965867]
36. Herten N, Otto T, Adolph D, Pause BM, Kumsta R, Wolf OT. Enhanced startle responsivity 24 hours after acute stress exposure. Behav Neurosci 2016 Oct;130(5):521-530. [doi: 10.1037/bne0000156] [Medline: 27362707]
37. Herten N, Pomrehn D, Wolf OT. Memory for objects and startle responsivity in the immediate aftermath of exposure to the Trier Social Stress Test. Behav Brain Res 2017 May 30;326:272-280. [doi: 10.1016/j.bbr.2017.03.002] [Medline: 28315373]
38. Wolf OT, Schulte JM, Drimalla H, Hamacher-Dang TC, Knoch D, Dziobek I. Enhanced emotional empathy after psychosocial stress in young healthy men. Stress 2015;18(6):631-637. [doi: 10.3109/10253890.2015.1078787] [Medline: 26365125]

39. Schmider E, Ziegler M, Danay E, Beyer L, Bühner M. Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. Methodology 2010 Jan;6(4):147-151. [doi: 10.1027/1614-2241/a000016]
40. Lumley T, Diehr P, Emerson S, Chen L. The importance of the normality assumption in large public health data sets. Annu Rev Public Health 2002;23:151-169. [doi: 10.1146/annurev.publhealth.23.100901.140546] [Medline: 11910059]
41. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. Introduction to Meta-Analysis. Hoboken, NJ, USA: John Wiley & Sons; 2009.
42. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol 2013 Nov 26;4:863 [FREE Full text] [doi: 10.3389/fpsyg.2013.00863] [Medline: 24324449]
43. Hawn SE, Paul L, Thomas S, Miller S, Amstadter AB. Stress reactivity to an electronic version of the Trier Social Stress Test: a pilot study. Front Psychol 2015 Mar 29;6:724 [FREE Full text] [doi: 10.3389/fpsyg.2015.00724] [Medline: 26074862]
44. Montero-López E, Santos-Ruiz A, García-Ríos MC, Rodríguez-Blázquez R, Pérez-García M, Peralta-Ramírez MI. A virtual reality approach to the Trier Social Stress Test: contrasting two distinct protocols. Behav Res Methods 2016 Mar;48(1):223-232. [doi: 10.3758/s13428-015-0565-4] [Medline: 25673321]
45. Fich LB, Jönsson P, Kirkegaard PH, Wallergård M, Garde AH, Hansen Å. Can architectural design alter the physiological reaction to psychosocial stress? A virtual TSST experiment. Physiol Behav 2014 Aug;135:91-97. [doi: 10.1016/j.physbeh.2014.05.034] [Medline: 24907691]
46. Shiban Y, Diemer J, Brandl S, Zack R, Mühlberger A, Wüst S. Trier Social Stress Test in vivo and in virtual reality: dissociation of response domains. Int J Psychophysiol 2016 Dec;110:47-55. [doi: 10.1016/j.ijpsycho.2016.10.008] [Medline: 27742258]
47. Jönsson P, Wallergård M, Osterberg K, Hansen AM, Johansson G, Karlson B. Cardiovascular and cortisol reactivity and habituation to a virtual reality version of the Trier Social Stress Test: a pilot study. Psychoneuroendocrinology 2010 Oct;35(9):1397-1403. [doi: 10.1016/j.psyneuen.2010.04.003] [Medline: 20451329]
48. Turner-Cobb JM, Asif M, Turner JE, Bevan C, Fraser DS. Use of a non-human robot audience to induce stress reactivity in human participants. Comput Human Behav 2019 Oct;99:76-85. [doi: 10.1016/j.chb.2019.05.019]
49. Zimmer P, Buttlar B, Halbeisen G, Walther E, Domes G. Virtually stressed? A refined virtual reality adaptation of the Trier Social Stress Test (TSST) induces robust endocrine responses. Psychoneuroendocrinology 2019 Mar;101:186-192. [doi: 10.1016/j.psyneuen.2018.11.010] [Medline: 30469086]
50. Dierolf AM, Schoofs D, Hessas EM, Falkenstein M, Otto T, Paul M, et al. Good to be stressed? Improved response inhibition and error processing after acute stress in young and older men. Neuropsychologia 2018 Oct;119:434-447. [doi: 10.1016/j.neuropsychologia.2018.08.020] [Medline: 30171874]
51. Merz CJ, Wolf OT, Schweckendiek J, Klucken T, Vaitl D, Stark R. Stress differentially affects fear conditioning in men and women. Psychoneuroendocrinology 2013 Nov;38(11):2529-2541. [doi: 10.1016/j.psyneuen.2013.05.015] [Medline: 23790683]
52. Sumter SR, Bokhorst CL, Miers AC, Van Pelt J, Westenberg PM. Age and puberty differences in stress responses during a public speaking task: do adolescents grow more sensitive to social evaluation? Psychoneuroendocrinology 2010 Nov;35(10):1510-1516. [doi: 10.1016/j.psyneuen.2010.05.004] [Medline: 20541871]
53. Westenberg PM, Bokhorst CL, Miers AC, Sumter SR, Kallen VL, van Pelt J, et al. A prepared speech in front of a pre-recorded audience: subjective, physiological, and neuroendocrine responses to the Leiden Public Speaking Task. Biol Psychol 2009 Oct;82(2):116-124. [doi: 10.1016/j.biopsycho.2009.06.005] [Medline: 19576261]
54. Narvaez Linares NF, Charron V, Ouimet AJ, Labelle PR, Plamondon H. A systematic review of the Trier Social Stress Test methodology: issues in promoting study comparison and replicable research. Neurobiol Stress 2020 Nov;13:100235 [FREE Full text] [doi: 10.1016/j.ynstr.2020.100235] [Medline: 33344691]
55. Campbell J, Ehlert U. Acute psychosocial stress: does the emotional stress response correspond with physiological responses? Psychoneuroendocrinology 2012 Aug;37(8):1111-1134. [doi: 10.1016/j.psyneuen.2011.12.010] [Medline: 22260938]
56. Gotthardt U, Schweiger U, Fahrenberg J, Lauer CJ, Holsboer F, Heuser I. Cortisol, ACTH, and cardiovascular response to a cognitive challenge paradigm in aging and depression. Am J Physiol 1995 Apr;268(4 Pt 2):R865-R873. [doi: 10.1152/ajpregu.1995.268.4.R865] [Medline: 7733395]
57. Neupert SD, Miller LM, Lachman ME. Physiological reactivity to cognitive stressors: variations by age and socioeconomic status. Int J Aging Hum Dev 2006;62(3):221-235. [doi: 10.2190/17DU-21AA-5HUK-7UFG] [Medline: 16625938]
58. Seeman TE, Robbins RJ. Aging and hypothalamic-pituitary-adrenal response to challenge in humans. Endocr Rev 1994 Apr;15(2):233-260. [doi: 10.1210/edrv-15-2-233] [Medline: 8026389]
59. Nicolson N, Storms C, Ponds R, Sulon J. Salivary cortisol levels and stress reactivity in human aging. J Gerontol A Biol Sci Med Sci 1997 Mar;52(2):M68-M75. [doi: 10.1093/gerona/52a.2.m68] [Medline: 9060972]
60. Hargittai E, Piper AM, Morris MR. From Internet access to Internet skills: digital inequality among older adults. Univ Access Inf Soc 2018 May 3;18(4):881-890. [doi: 10.1007/s10209-018-0617-5]
61. Yaribeygi H, Panahi Y, Sahraei H, Johnston TP, Sahebkar A. The impact of stress on body function: a review. EXCLI J 2017 Jul 21;16:1057-1072 [FREE Full text] [doi: 10.17179/excli2017-480] [Medline: 28900385]

62. Rimmele U, Zellweger BC, Marti B, Seiler R, Mohiyeddini C, Ehlert U, et al. Trained men show lower cortisol, heart rate and psychological responses to psychosocial stress compared with untrained men. Psychoneuroendocrinology 2007 Jul;32(6):627-635. [doi: 10.1016/j.psyneuen.2007.04.005] [Medline: 17560731]
63. Schlotz W, Kumsta R, Layes I, Entringer S, Jones A, Wüst S. Covariance between psychological and endocrine responses to pharmacological challenge and psychosocial stress: a question of timing. Psychosom Med 2008 Sep;70(7):787-796. [doi: 10.1097/PSY.0b013e3181810658] [Medline: 18725434]
64. Domes G, Heinrichs M, Rimmele U, Reichwald U, Hautzinger M. Acute stress impairs recognition for positive words--association with stress-induced cortisol secretion. Stress 2004 Sep;7(3):173-181. [doi: 10.1080/10253890412331273213] [Medline: 15764014]
65. Hellhammer J, Schubert M. The physiological response to Trier Social Stress Test relates to subjective measures of stress during but not before or after the test. Psychoneuroendocrinology 2012 Jan;37(1):119-124. [doi: 10.1016/j.psyneuen.2011.05.012] [Medline: 21689890]
66. Seo EJ, Ahn JA, Hayman LL, Kim CJ. The association between perceived stress and quality of life in university students: the parallel mediating role of depressive symptoms and health-promoting behaviors. Asian Nurs Res (Korean Soc Nurs Sci) 2018 Sep;12(3):190-196 [FREE Full text] [doi: 10.1016/j.anr.2018.08.001] [Medline: 30103040]
67. Strizhitskaya O, Petrash M, Savenysheva S, Murtazina I, Golovey L. Perceived stress and psychological well-being: the role of the emotional stability. In: Proceedings of the 7th Annual International Conference on Cognitive - Social, and Behavioural Sciences. 2018 Presented at: icCSBs '18; November 12-14, 2018; Moscow, Russia p. 162. [doi: 10.15405/epsbs.2019.02.02.18]
68. Bolier L, Haverman M, Westerhof GJ, Riper H, Smit F, Bohlmeijer E. Positive psychology interventions: a meta-analysis of randomized controlled studies. BMC Public Health 2013 Feb 08;13:119 [FREE Full text] [doi: 10.1186/1471-2458-13-119] [Medline: 23390882]
69. Weiss LA, Westerhof GJ, Bohlmeijer ET. Can we increase psychological well-being? The effects of interventions on psychological well-being: a meta-analysis of randomized controlled trials. PLoS One 2016 Jun 21;11(6):e0158092 [FREE Full text] [doi: 10.1371/journal.pone.0158092] [Medline: 27328124]
70. Andrews J, Ali N, Pruessner JC. Reflections on the interaction of psychogenic stress systems in humans: the stress coherence/compensation model. Psychoneuroendocrinology 2013 Jul;38(7):947-961. [doi: 10.1016/j.psyneuen.2013.02.010] [Medline: 23522990]
71. McHugh RK, Kneeland ET, Edwards RR, Jamison R, Weiss RD. Pain catastrophizing and distress intolerance: prediction of pain and emotional stress reactivity. J Behav Med 2020 Aug;43(4):623-629 [FREE Full text] [doi: 10.1007/s10865-019-00086-5] [Medline: 31376099]
72. Powell N, Tarr A, Sheridan J. Psychosocial stress and inflammation in cancer. Brain Behav Immun 2013 Mar;30 Suppl:S41-S47. [doi: 10.1016/j.bbi.2012.06.015] [Medline: 22790082]
73. Travis F, Haaga DA, Hagelin J, Tanner M, Nidich S, Gaylord-King C, et al. Effects of Transcendental Meditation practice on brain functioning and stress reactivity in college students. Int J Psychophysiol 2009 Feb;71(2):170-176. [doi: 10.1016/j.ijpsycho.2008.09.007] [Medline: 18854202]
74. Limm H, Angerer P, Heinmueller M, Marten-Mittag B, Nater UM, Guendel H. Self-perceived stress reactivity is an indicator of psychosocial impairment at the workplace. BMC Public Health 2010 May 14;10:252 [FREE Full text] [doi: 10.1186/1471-2458-10-252] [Medline: 20470413]
75. Ostberg M. Parental stress, psychosocial problems and responsiveness in help-seeking parents with small (2-45 months old) children. Acta Paediatr 1998 Jan;87(1):69-76. [doi: 10.1080/08035259850157903] [Medline: 9510451]
76. Kirschbaum C. Why we need an online version of the Trier Social Stress Test. Psychoneuroendocrinology 2021 Mar;125:105129. [doi: 10.1016/j.psyneuen.2021.105129] [Medline: 33493741]
77. Goodday SM, Friend S. Unlocking stress and forecasting its consequences with digital technology. NPJ Digit Med 2019 Jul 31;2:75 [FREE Full text] [doi: 10.1038/s41746-019-0151-8] [Medline: 31372508]
78. Baird A, Triantafyllopoulos A, Zänkert S, Ottl S, Christ L, Stappen L, et al. An evaluation of speech-based recognition of emotional and physiological markers of stress. Front Comput Sci 2021 Dec 6;3:750284 [FREE Full text] [doi: 10.3389/fcomp.2021.750284]
79. Ghadiyaram D, Tran D, Mahajan D. Large-scale weakly-supervised pre-training for video action recognition. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019 Jun Presented at: CVPR '19; June 15-20, 2019; Long Beach, CA, USA p. 12038-12047 URL: https://ieeexplore.ieee.org/document/8953267/ [doi: 10.1109/cvpr.2019.01232]
80. Golden EA, Zweig M, Danieletto M, Landell K, Nadkarni G, Bottinger E, et al. A resilience-building app to support the mental health of health care workers in the COVID-19 era: design process, distribution, and evaluation. JMIR Form Res 2021 May 05;5(5):e26590 [FREE Full text] [doi: 10.2196/26590] [Medline: 33872189]

Abbreviations

API: application programming interface

C-DST: Control - Digital Stress Test
DST: Digital Stress Test
MIST: Montreal Imaging Stress Task
PANAS: Positive and Negative Affect Schedule
TSST: Trier Social Stress Test
VAS: Visual Analogue Scale

Edited by A Mavragani; submitted 28.07.21; peer-reviewed by R Ciorap, P Bhadra, H Mehdizadeh; comments to author 08.03.22; revised version received 24.05.22; accepted 26.05.22; published 15.07.22.

Please cite as:
Norden M, Hofmann AG, Meier M, Balzer F, Wolf OT, Böttinger E, Drimalla H
Inducing and Recording Acute Stress Responses on a Large Scale With the Digital Stress Test (DST): Development and Evaluation Study
J Med Internet Res 2022;24(7):e32280
URL: https://www.jmir.org/2022/7/e32280
doi:10.2196/32280
PMID:35838765

Original Paper

Emotional and Physical Health Impact in Children and Adolescents and Their Caregivers Using Open-source Automated Insulin Delivery: Qualitative Analysis of Lived Experiences

Katarina Braune[1,2,3], MD; Niklas Krug[1]; Christine Knoll[1,2,4], MD; Hanne Ballhausen[1,2,5], MSc; Axel Thieffry[6,7], PhD; Yanbing Chen[8], PhD; Shane O'Donnell[4], PhD; Klemens Raile[1], MD; Bryan Cleal[9], PhD

[1]Department of Paediatric Endocrinology and Diabetes, Charité - Universitätsmedizin Berlin, Berlin, Germany

[2]Berlin Institute of Health, Berlin, Germany

[3]Institute of Medical Informatics, Charité - Universitätsmedizin Berlin, Berlin, Germany

[4]School of Sociology, University College Dublin, Dublin, Ireland

[5]#dedoc° Diabetes Online Community, Berlin, Germany

[6]Jay Keasling Faculty, BioInnovation Institute, Center for Biosustainability, Technical University of Denmark, Copenhagen, Denmark

[7]Intomics A/S, Kongens Lyngby, Denmark

[8]School of Public Health, Physiotherapy & Sports Science, University College Dublin, Belfield, Ireland

[9]Diabetes Management Research, Steno Diabetes Center Copenhagen, Herlev, Denmark

Corresponding Author:
Katarina Braune, MD
Department of Paediatric Endocrinology and Diabetes
Charité - Universitätsmedizin Berlin
Augustenburger Platz 1
Berlin, 13353
Germany
Phone: 49 30450616454
Email: katarina.braune@charite.de

Abstract

Background: Given the limitations in the access and license status of commercially developed automated insulin delivery (AID) systems, open-source AID systems are becoming increasingly popular among people with diabetes, including children and adolescents.

Objective: This study aimed to investigate the lived experiences and physical and emotional health implications of children and their caregivers following the initiation of open-source AID, their perceived challenges, and sources of support, which have not been explored in the existing literature.

Methods: Data were collected through 2 sets of open-ended questions from a web-based multinational survey of 60 families from 16 countries. The narratives were thematically analyzed, and a coding framework was identified through iterative alignment.

Results: A range of emotions and improvements in quality of life and physical health were reported, as open-source AID enabled families to shift their focus away from diabetes therapy. Caregivers were less worried about hypoglycemia at night and outside their family homes, leading to increased autonomy for the child. Simultaneously, the glycemic outcomes and sleep quality of both the children and caregivers improved. Nonetheless, the acquisition of suitable hardware and technical setup could be challenging. The #WeAreNotWaiting community was the primary source of practical and emotional support.

Conclusions: Our findings show the benefits and transformative impact of open-source AID and peer support on children with diabetes and their caregivers and families, where commercial AID systems are not available or suitable. Further efforts are required to improve the effectiveness and usability and facilitate access for children with diabetes, worldwide, to benefit from this innovative treatment.

International Registered Report Identifier (IRRID): RR2-10.2196/15368

(J Med Internet Res 2022;24(7):e37120) doi:10.2196/37120

KEYWORDS

automated insulin delivery; closed-loop; do-it-yourself; open source; peer support; patient-reported outcomes; lived experiences; qualitative analysis; mobile phone

Introduction

Background

Type 1 diabetes (T1D) is a challenging chronic condition for children, adolescents, and their caregivers and is associated with long-term macro- and microvascular complications and the consequent risk of increased morbidity and mortality. Therapeutic guidelines of the International Society for Pediatric and Adolescent Diabetes recommend a target hemoglobin A_{1c} (HbA_{1c}) level of <7% for children and adolescents with T1D, albeit a target that must be balanced with the individual disease burden and risk of hypoglycemia [1].

The management of diabetes is particularly challenging during childhood and adolescence. Day-to-day tasks often involve an entire family. Children show variability in insulin sensitivity related to physical growth and sexual maturation, which requires frequent adjustments in insulin dosing [2]. With the dynamic physical activity and nutritional intake of young children, their glycemic levels can fluctuate rapidly [3]. In addition, the transition of responsibility for diabetes management from caregivers to children and their increasing independence during adolescence can often further complicate this difficult dynamic. Adolescents and young adults with diabetes frequently struggle to meet the recommended glycemic targets and are particularly vulnerable to acute complications, such as severe hypoglycemia and diabetic ketoacidosis [4,5]. Living with T1D also impacts the quality of life and mental health [6]. Thus, psychosocial support and individualized treatment play an important role in diabetes care in this age group [1].

Recent advances in diabetes technology have led to the development of automated insulin delivery (AID) systems, also known as hybrid closed-loop, closed-loop, or artificial pancreas systems. In AID, a control algorithm automatically adjusts the insulin delivery of an insulin pump in response to readings from a continuous glucose monitor (CGM) to help improve glycemic levels and variability and reduce the day-to-day burden in diabetes management [7-10].

Although commercially developed AID systems have recently become available in select countries, not all are licensed for use by children. Currently, CamAPS FX (CamDiab Ltd) is the only AID system that has received regulatory approval for children aged ≤7 years but is restricted in interoperability and only compatible with one specific CGM and pump model, only available in select European countries, and must be individually purchased on a subscription basis. Young children are often the last cohort to be included in a clinical trial. Off-label use of commercial AID in this group shows only minor time in range (TIR) and HbA_{1c} improvements compared with older individuals, indicating a higher hypoglycemia risk for this age group [11,12].

Parents and caregivers of children with diabetes have been in the vanguard of the drive toward AID systems. Under the hashtag *#WeAreNotWaiting*, a web-based patient community has sought to create resources and tools for diabetes management since 2013. The movement began with the “Nightscout” project, where caregivers created a cloud-based platform for alerts and remote glucose monitoring for their children. Eventually, the community developed control algorithms for the AID. In these “do-it-yourself” or “open-source” AID systems, commercially available sensors for CGM and insulin pumps are linked to a microcontroller or an app on a smartphone. The source code and documentation of these systems were shared freely on the web. In addition, the community provides both practical and emotional peer support with setup and maintenance. To date, open-source AID systems have not been approved by regulatory bodies and must be built and used at an individual risk. An estimated number of >10,000 individuals, worldwide, use open-source AID. Approximately 20% of these users are children and adolescents, where their caregivers are building and maintaining the systems on their behalf [13,14].

Although evidence based on the clinical outcomes of open-source AID is growing, there are relatively few published studies on the lived experiences of people with diabetes using this technology and fewer still, concerning children and adolescents with diabetes and their caregivers. Previous studies have found improvements in HbA_{1c} and percentage TIR in various age groups, including children and adolescents [13-16]. As part of the Outcomes of Patients’ Evidence with Novel, Do-it-Yourself Artificial Pancreas Technology (OPEN) [17] project, we previously assessed self- or caregiver-reported clinical outcomes [14,15] and motivations [14] to build open-source AID. Improved sleep quality was a primary reason for caregivers to use AID, followed by improved glycemia and reduced complication risk for the child, and the option of remote monitoring and control via the internet, thus reducing disease burden and enabling more independence for children. A recently published international consensus statement on open-source AID supported its use for children and adolescents, as long as the child's welfare is being considered by health care professionals (HCPs) and caregivers who are setting up open-source AID systems for their children, with the child's assent and engagement [18].

Objectives

This study aimed to examine four specific, albeit interrelated, aspects of the lived experiences of children and adolescents with diabetes and their caregivers on their journey to becoming open-source AID users: (1) the emotional health implications of open-source AID, (2) the experience of changes to physical health using open-source AID, (3) perceived challenges with the implementation and maintenance of open-source AID, and (4) sources of support during the implementation and maintenance of open-source AID. Self-reported glycemic outcomes and sleep have also been reported to provide further context for lived experience data.

Methods

The results were obtained from answers to 2 open-ended questions included in a cross-sectional web-based survey examining the use of open-source AID. The survey titled "DIWHY" was conducted between November 2018 and March 2019 [17].

Ethics Approval

Ethics approval was provided by Charité—Universitätsmedizin Berlin (EA2/140/18).

Survey Design

The survey (Multimedia Appendix 1) was created by the patient-led OPEN consortium in collaboration with further users of open-source AID and was piloted with a small group before the final release. The Checklist for Reporting Results of Internet E-Surveys (CHERRIES) guideline was used to guide the survey development [19]. The survey included 39 items in total, including questions on the child's demographic information, the open-source AID system in use, HbA_{1c}, and TIR before and after initiation, and 2 composite open-ended questions, which sought to capture lived experiences with open-source AID in the form of narratives. Participants could enter a free-text answer with up to 1000 words for each of the 2 questions.

The first question inquired about the individual journey of the caregiver and child toward setting up and using an open-source AID system, including sources of information, support, motivation, and emotional impact:

> *If you would like, please share your personal story about why you decided to build your own artificial pancreas system and how you got started. Feel free to share any experiences that had a significant impact on how you manage your diabetes as well. This story can be as short or as long as you wish.*

The second question addressed the perceived changes following the initiation of open-source AID and the challenges experienced:

> *When reflecting on your personal DIY closed-loop story, you may want to consider the following: When did you first hear about DIY closed-loop systems and how did you look for further information? Were there any key events or experiences that were a factor in your decision to begin closed looping? Was there anyone else involved in helping you come to decision to begin DIY closed-looping? For example a friend, family member or an online support group? What were your emotions in the lead-up to building your DIY closed-loop system? For example, had you any major hopes or fears?*

Participants and Recruitment

Participants were eligible if they were caregivers of a child or adolescent with diabetes (type 1, 2, or other), using an open-source AID. There were no restrictions on age, time since diagnosis, or commencement of open-source AID.

Participants were recruited using public announcements on the OPEN project website and social media channels, such as the Facebook groups "Looped" (>6000 members) and "AndroidAPS Users" (>1800 members as of November 2018), regional subgroups, and tweets under the hashtags *#WeAreNotWaiting* and *#DIYAPS*. All posts were organic, meaning that their web-based reach was not affected by any monetary influence. Participants consented electronically and joined voluntarily and anonymously with the children's assent. The survey was available in German and English.

Data Collection and Analysis

Data were collected and managed using the REDCap (Research Electronic Data Capture; Vanderbilt University) electronic data capture tools hosted at Charité. Following deidentification of the data set, qualitative analysis was performed using NVivo 12 (QSR International, 2018) software. The narratives were analyzed using an approach based on the principles of Template Analysis [20]. Acknowledging the response priming included in the framing of the open-ended questions, initial coding (by KB, CK, and NK) sorted the data in accordance with 4 predefined topics: physical health impact, emotional impact, sources of support, and perceived challenges. To establish alignment, all 3 coders analyzed and sorted the first 30 narratives into 4 topics. Using the "coding comparison" function in NVivo, it was established that there was a high level of agreement among the coders. Level of agreement was defined as the number of units of agreement divided by the total units of measure within the data item, as a percentage. In the next phase of data analysis, all data extracts sorted into the 4 topics were coded inductively and independently by the 3 coders, which led to an extensive set of descriptive codes. Finally, codes were collaboratively collated and used to establish a set of higher-level codes, each of which was described in detail in a codebook. The template codebook was refined and modified in discussions between the 3 coders and the project group.

To test the utility and resonance of the themes as captured in the template, 2 coders (HB and BC) further used the template to analyze the narratives independently of one another. The initial group of coders (KB, CK, and NK) then refined the template based on the coding and feedback provided during this process. A third independent coder (SO) then analyzed the data using the refined template. After this final review of coded responses and the template, it was agreed that code saturation had been achieved, and all major themes were identified.

Retrospective and caregiver-reported clinical outcome data were analyzed within the R (R Foundation for Statistical Computing) programming framework. Only respondents who reported at least one value before and after open-source AID commencement were considered, leading to sample sizes of N=52 and N=36 for HbA_{1c} and TIR, respectively. The HbA_{1c} values were averaged. Moreover, 1-tailed Student *t*tests were conducted with the parameter *paired=TRUE*. Figures were produced using the ggplot2 package.

Results

Participant Characteristics

In total, 60 caregivers (35.7% of all 168 participants in the DIWHY study) from 16 countries responded to the open-ended questions on behalf of their children, and there were combined 107 responses to both questions. All children and adolescents were diagnosed with T1D, aged between 3 and 20 years, and using an open-source AID for a duration of <1 month and up to 3 years. The caregiver and child demographics as well as the clinical features of the 60 participants who responded to the open-ended questions are summarized in Table 1, whereas the characteristics of the other 108 participants of the DIWHY study are included in Table S1 in Multimedia Appendix 2.

Of the 60 children and adolescents, the average HbA_{1c} levels (of participants with reported measures both before and after AID commencement, see Methods section) decreased from 7.0% (SD 0.8; 53 mmol/mol) to 6.3% (SD 0.7; 45 mmol/mol; 1-tailed paired *t* test; $P<.001$; Figure 1), and TIR increased from 60.7% (SD 15.1) to 80.4% (SD 9.1; 1-tailed paired *t* test; $P<.001$) following the initiation of open-source AID (Figure 2).

Table 1. Children's and caregivers' demographic and self-reported clinical characteristics (N=60).

	Children and adolescents
Child's gender, n (%)	
Female	26 (43)
Male	34 (57)
Other	0 (0)
Child's age (years), mean (SD)	10.0 (4.5)
Type of diabetes, n (%)	
Type 1	60 (100)
Type 2	0 (0)
Other	0 (0)
Duration of diabetes (years), mean (SD)	5.3 (4.3)
Duration of open-source AID[a] use (months), mean (SD)	10.9 (9.2)
Type of open-source AID, n (%)	
AndroidAPS	28 (47)
OpenAPS	21 (35)
Loop	17 (28)
Other[b]	2 (3)
Region and country of residence, n (%)	
Europe	47 (78)
Germany	12 (20)
United Kingdom	9 (15)
Finland	7 (12)
Sweden	5 (8)
Czech Republic	3 (5)
Spain	3 (5)
Slovakia	3 (5)
Others[c]	14 (12)
North America	6 (10)
United States	4 (7)
Canada	2 (3)
Asia	
South Korea	3 (5)
Western Pacific	
Australia	5 (8)
Caregiver's education: highest completed, n (%)	
University degree or diploma	38 (63)
Doctorate	9 (15)
High school	8 (13)
Other	5 (8)
Caregiver's occupational status, n (%)	
Full-time	39 (65)
Part-time	15 (25)

	Children and adolescents
Unemployed	3 (5)
Other	3 (5)
Annual household income (US $), n (%)	
<20,000	4 (7)
20,000 to 34,999	5 (8)
35,000 to 49,999	4 (7)
50,000 to 74,999	11 (18)
75,000 to 99,999	10 (17)
>100,000	16 (27)
I would rather not say	3 (3)
Not stated	7 (12)

[a]AID: automated insulin delivery.

[b]"Open loop with AndroidAPS" and "custom development."

[c]Austria, Bulgaria, Croatia, and Greece.

Figure 1. Outcomes of open-source automated insulin delivery (AID) implementation. Density distributions of hemoglobin A1c (HbA1c) before and after commencement of open-source AID (line colors); n=52.

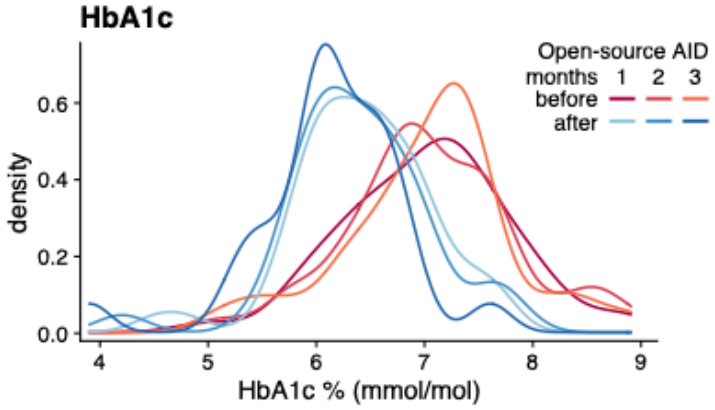
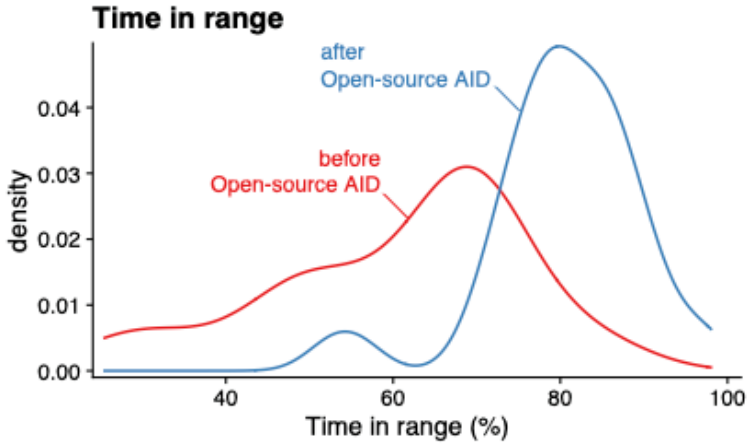


Figure 2. Outcomes of open-source automated insulin delivery (AID) implementation. Density distributions of time in range (70-180 mg/dL/3.9-10.0 mmol/L) before and after commencement of open-source AID (line colors); n=36.



Template Analysis

Overview

A total of 4 topics, “Emotional and Quality of Life Impact,” “Physical Health Impact,” “Challenges,” and “Support” were used to organize the qualitative data, recognizing the fact that participants’ responses were partially primed by the framing of the open-ended questions. The data were subsequently analyzed to generate codes within these topics to expand and illustrate them. The codes are described with examples of illustrative quotes, the number of occurrences, and the respondents’ profiles, as shown in Table 2.

Table 2. Final codebook template including deductively (A-D) and inductively (A1-D2) developed codes.

Topic	Occurrences[a], n	Illustrative quote	Respondent profile
Emotional and quality of life impact (A)			
Worry and fear (A1): describes difficult emotions such as worry and fear of caregivers related to living with and managing diabetes, experiencing hypoglycemia, and developing long-term complications. It also refers to the concern of not being able to build and maintain the open-source AID[b].	16	"I was very skeptical and scared. Over time more information became available and the safety became clear and compelling. We realized we would be safer with a Loop than without. I was scared that others would not be able to comprehend this (because even endocrinologists fail to understand fully the burden and dynamism of type 1) and that they would question whether we were putting our child at risk and make a report about us to child well-being authorities."	Caregiver of a boy aged 13 years, from Australia; aged 6 years at diagnosis; using Loop for 2.5 years
Desperation and frustration (A2): describes feelings of desperation and frustration of caregivers related to living with diabetes and caring for a child with diabetes, diabetes management, and the implementation of the open-source AID.	14	"As a mom I was desperate, I was tired from being up all night, I was getting frustrated from teen hormones and I was willing to try almost anything to help both of us."	Caregiver of a girl aged 17 years, from the United States; aged 2 years at diagnosis; using Loop for 33 months
Uncertainty (A3): describes uncertainty and insecurities of caregivers regarding legal grounds, missing regulatory guidelines, and the trust of reliability in an open-source AID system.	6	"Nevertheless, there is still a legal uncertainty and at the moment we just dare to use the loop in our own four walls. In the morning we switch to the normal AnyDana A app, in the evening back to AndroidAPS."	Caregiver of a boy aged 12 years, from Germany; aged 11 years at diagnosis; using OpenAPS for 2 weeks
Anticipation, hope, and wishes (A4): describes positive and hopeful emotional states of anticipation and great expectations of caregivers that lie on the AID for improved diabetes management and hope for improved quality of life. Also includes wishes for access to an AID system for everyone.	24	"Major driver for the project was to give my son more years without complications by lowering the HbA_{1c}[c]."	Caregiver of a boy aged 18 years, from Finland; aged 1 year at diagnosis; using OpenAPS for 1 year
Excitement, appreciation, and satisfaction (A5): describes all positive emotions of caregivers and children related to the experience with the open-source AID in daily use including excitement, happiness, satisfaction with the results, and appreciation.	51	"I remember the exact place I stood watching the [OpenAPS] log [roll] and seeing the [preflight] was successful and then that the loop was complete. I was in shock that we could do this and that I could afford it and that my child was going to [be] better off because of this. It was a defining moment in my life as a parent. No one could stop me giving my child the care they needed anymore. Especially not a company who places shareholders above clients (which legally they must do). I was no longer at the mercy of markets, profits, politics and whims, I had the capacity to provide for my child again."	Caregiver of a boy aged 13 years, from Australia; aged 6 years at diagnosis; using Loop for 2.5 years
Security and reassurance (A6): relates to caregivers feeling more empowered, more secure, and reassured owing to the use of an open-source AID system, through automation, remote monitoring, and control, as well as experiencing success and observing the success of others using an open-source AID.	45	"Our child never woke up if she had a low even though her pump was sounding a very loud alarm. And because she slept in her own room we were afraid of sometimes not hearing the pump alarm either. So Nightscout sounded like the perfect solution, as we could then be woken up by any mobile phone or iPad. This added a lot to our feeling of security."	Caregiver of a girl aged 10 years, from Finland; aged 7 years at diagnosis; using OpenAPS for 3 months

Topic	Occurrences[a], n	Illustrative quote	Respondent profile
Child empowerment and independence (A7): describes the degree of independence, autonomy, and self-determination in children and adolescents using the open-source AID, enabling them to participate in daily life and social activities in a responsible and self-determined way.	25	"Daughter can work without having to phone me for advice. She has been on holiday [for the] first time without parents. She[...] now feels confident to consider leaving home."	Caregiver of a girl aged 20 years, from Croatia; aged 10 years at diagnosis; using OpenAPS for 3 months
Physical health impact (B)			
Glycemic outcome improvement (B1): refers to improved time in range and HbA_{1c} levels, less glucose variability, fewer hypo- and hyperglycemic events, and reduced long-term complication risk.	36	"Every single morning she's in range. If at night she's not, we know that by the morning she will be, and she [will get] there safely. It's really good."	Caregiver of a girl aged 18 years, from the United Kingdom; aged 11 years at diagnosis; using AndroidAPS for 8 months
Quality of life improvement (B2): refers to the mentioned improvements of quality of life and describes the degree to which an individual is healthy, comfortable, and able to participate in or enjoy life events.	14	"I keep a continuous discussion with my twins that both use DIY closed loops, through texting. I use this way to share my remote observations on their status, while they concentrate on their university studies, or simply enjoy their lives. I inform them this way about a failing connection, a reservoir getting empty, a battery needing charging, or to drink some juice to avoid a coming low."	Caregiver of a boy, aged 20 years, from Greece; aged 2 years at diagnosis; using OpenAPS for 1 year
Improved sleep (B3): denotes all aspects of improved sleep quality for either caregivers or children such as increased sleep duration, fewer sleep interruptions, and feeling better rested in the morning.	40	"It's been as good as expected, and better still as now we sleep. You forget how much sleep deprivation clouds your judgment."	Caregiver of a boy aged 8 years, from the United Kingdom; aged 7 years at diagnosis; using AndroidAPS for 3 months
Facilitated diabetes management (B4): relates to the simplifications of the individual diabetes management due to the open-source AID, such as fewer interactions with the technology or between caregiver and child; for example, through remote control and automation. It also includes the age-appropriate transfer of responsibilities from caregivers to adolescents to self-manage diabetes therapy.	42	"There is no comparison with earlier. There used to be 5-6 blood measurements per child per day, and that was all. With or without a pump, every meal was a challenge. For 1.5 years, the APS has been adjusting the blood sugar value after the bolus, adding more insulin if the value increases, or adjusting the delivery if the value drops."	Caregiver of a boy aged 20 years, from Greece; aged 1 year at diagnosis; using OpenAPS for 17 months
Challenges (C)			
Access to technology (C1): relates to the issue concerning obtaining access to the component parts of an open-source AID system, such as loopable pumps and supplies, CGM[d], and additional required hardware.	27	"Getting the hardware was most frustrating. I tried to buy the hardware from the manufacturer but in Sweden you could not do that without a subscription from your doctor. I ended up getting a second hand Dana R pump from another patient who upgraded to a newer pump."	Caregiver of a boy aged 3.5 years, from Sweden; aged 2 years at diagnosis; using AndroidAPS for 4 months
Out-of-pocket expenses (C2): describes barriers regarding out-of-pocket expenses and cost for the hardware and supplies related to insurance coverage, household income, and other financial challenges in access.	6	"We were concerned about the cost of sensors. They are not covered by private health here and it cost approximately US $5000 a year when we started. Now kids are covered, but when they turn 21 that ends. We are still worried about covering that bill in the future."	Caregiver of a boy aged 13 years, from Australia; aged 6 years at diagnosis; using Loop for 2.5 years
Self-perceived lack of technical skills (C3): denotes the issue of yet self-perceived limited knowledge and missing technical skills caregivers are experiencing to set up open-source AID initially.	9	"Major fears I wouldn't be able to understand the technology."	Caregiver of a girl aged 12 years, from Australia; aged 11 years at diagnosis; using Loop for 1 month

Topic	Occurrences[a], n	Illustrative quote	Respondent profile
Lacking health care provider support (C4): relates to instances where caregivers reflect upon their children's health care providers' lack of support and negative attitudes.	14	"Fight with our own diabetologist to get a DANA RS prescribed. Although we didn't talk openly about looping, she has repeatedly emphasized that we only want the DANA RS pump for looping, which is not allowed. We have won, but now hide the loop, which cannot be a permanent state. We need medical care in which we can communicate openly."	Caregiver of a boy aged 12 years, from Germany; aged 11 years at diagnosis; using AndroidAPS for 2 weeks
Impracticability of carrying additional devices (C5): relates to the necessity for children and adolescents having to carry additional devices for open-source AID and protect them from breaking.	9	"It also meant that our daughter had to carry an extra item, i.e. the mini-computer, with her during the day."	Caregiver of a girl aged 10 years, from Finland; aged 4 years at diagnosis; using OpenAPS for 2 months
Transition from childhood to adulthood (C6): describes challenges associated with the transition from childhood to adulthood, regarding physical and hormone-related changes during puberty and psychosocial challenges in adolescents living with T1D[e] and taking over responsibility for their own therapy with an open-source AID.	10	"While our control has improved, it is still significantly more variable than I would expect based on the results I see from others in the community. My son is highly insulin sensitive [...], variable in his activity level and intensity [...], and experiencing substantial swings in carb ratios, basal rates, and insulin sensitivities as he is going through great physiological changes in puberty."	Caregiver of a boy aged 11 years, from the United States; aged 8 years at diagnosis; using OpenAPS for 1 year
Setup and maintenance effort (C7): relates to difficulties caregivers experience while setting up open-source AID. This includes an unexpected high time effort and multiple throwbacks while initially setting up the system, technical difficulties with running and maintaining the system, and fine-tuning to find the right settings and parameters to accomplish desired results.	54	"As a family, we feel very happy that we can finally control the blood sugar levels of our children in the desired area, even if it takes great care to do everything right. Batteries (pump, CGM, mobile phone, OpenAPS computer) must be regularly charged or exchanged, the CGM must be continuously calibrated, insulin must be refilled, every 3 days you exchange the catheter, every 14 days the CGM, etc. With such a result, no problem. The hundreds of hours I've spent on it are worth it."	Caregiver of a boy aged 20 years, from Greece; aged 1 year at diagnosis; using OpenAPS for 17 months
Support (D)			
Community peer support (D1): includes actively received or provided community peer support. This support could either be provided on the web through social media groups and communities or in person through life events, individual people, or meet-ups. Does not include individual key people or role models.	45	"So in that same Facebook group I started to learn about DIY artificial pa[n]creases and I joined another, international group called Looped to learn more. I then asked around and I was told that OpenAPS was the most advanced of the three options and decided to go for that."	Caregiver of a girl aged 10 years, from Finland; aged 7 years at diagnosis; using OpenAPS for 9 months
Individuals as role models (D2): describes one or multiple key people, often members of the #WeAreNotWaiting community, who inspired or directly supported caregivers and children in building an open-source AID.	15	"I found Tim Street's Diabettech website and started following him on twitter/blog at [the] same time. He was coming to speak at a medical conference in Edinburgh and was going to a [type 1] meet up. I gate-crashed the meet in the pub and had to wait until the end[...] I asked him to show me his pancreas! [...] Tim organized the first U.K. meet up in London and offered me an old transmitter which would complete my build. My son and I flew to London and we got going that evening."	Caregiver of a boy aged 12 years, from the United Kingdom; aged 8 years at diagnosis; using OpenAPS
Web-based resources (D3): describes web-based resources such as wiki blogs, tutorials, websites, webinars, and other documentation.	19	"Once I had the equipment, I set the system up in two nights, the instructions available on the web are very clear and I found it easier tha[n] expected."	Caregiver of a girl aged 10 years, from Finland aged 6 years at diagnosis; using OpenAPS for 3 months

Topic	Occurrences[a], n	Illustrative quote	Respondent profile
Health care professionals (D4): this code refers to the support provided by health care professionals, such as pediatricians, endocrinologists, and other members of the diabetes teams, including help with setup, access to components, and fine-tuning of settings.	8	"Endocrinologist was supportive even though legally couldn't recommend it."	Caregiver of a girl aged 12 years, from Australia; aged 11 years at diagnosis; using Loop for 1 month

[a]Defined by the number of codes assigned to a text segment.

[b]AID: automated insulin delivery.

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

[d]CGM: continuous glucose monitor.

[e]T1D: type 1 diabetes.

Topic 1: Emotional and Quality of Life Impact

For respondents, experiences with the initiation of open-source AID were associated with a range of emotions, from worry, despair, and great hopes before use, to excitement, relief, and a feeling of empowerment after implementing the system. Caregivers in the sample expressed concerns when opting to choose an open-source AID, but it also highlights the deep-rooted frustration and dissatisfaction with commercially available solutions for diabetes management. Therefore, choosing to opt for an open-source AID was never a decision taken lightly but at the point when all other options appeared inadequate and insufficient.

Once the choice was made, quality of life improvements and reductions in the burden of diabetes management were frequently mentioned. With the automation of insulin delivery, families could reboot everyday life without diabetes management being constantly the center of attention, empowering children and caregivers to experience more freedom and flexibility:

> *Now we plan for things in our lives. We have been thinking of getting a pet, [and have] started to remodel our house. [We] made sure both kids have passports because now it feels like we actually can travel and show them the world.* [Caregiver of boy aged 8 years, from Sweden; aged 1 year at diagnosis; using OpenAPS and AndroidAPS for 1.5 years]

The option to remotely follow and control glycemic levels, treatments, and insulin delivery via Nightscout reassured caregivers was specifically mentioned as a reason to choose open-source AID. Caregivers experienced fewer worries about their children experiencing hypoglycemia at night or away from home, which led to greater independence, empowerment, and age-appropriate participation of children in their own treatment.

The complexity of the implementation process of open-source AID raised concerns among some of the respondents, who were initially worried about not being able to manage the technical setup on their own. Uncertainties regarding the safety of new and unfamiliar therapies have also been mentioned. Furthermore, they were unsure whether the new treatment would be accepted by their children's health care team as well as their wider social environment. In addition, some expressed the need for regulatory approval and improved access to AID for everyone:

> *I wonder how it can be that such a development is not already established? Why does it take so long? Do the old systems have to be remunerated? The loopers show how it works, how can it be that with so much added value, the professional institutions are still so lethargic?* [Caregiver of a boy, from Germany, aged 1 year at diagnosis; using Loop for 3 months]

Overall, caregivers described the transition to open-source AID as a predominantly positive experience for the entire family. They were highly satisfied with the outcomes and benefits for their children's emotional and physical health and perceived open-source AID as the best therapy option available:

> *If I could give my pancreas to my son, I would. This is the next best available option.* [Caregiver of a boy aged 12 years, from the United Kingdom; aged 3 years at diagnosis; using Loop for 1 month]

Topic 2: Physical Health Impact

Improvements in glycemia, such as improvements in HbA_{1c} and TIR levels, as well as less hypoglycemia and fewer glucose fluctuations, have been extensively described:

> *Every single morning she's in range. If at night she's not we know that by the morning she will be, and she [will get] there safely.* [Caregiver of a girl aged 18 years, from the United Kingdom; aged 11 years at diagnosis; using AndroidAPS for 8 months]

In addition to diabetes-related health improvements, better sleep quality was frequently highlighted by the respondents. Before using an open-source AID, many caregivers were not able to sleep through the night as they were concerned with nighttime hypoglycemia or the administration of correction doses of insulin, poor sleep, and reduced quality of life. With an open-source AID, they were released from frequent check-ups and the associated emotional pressure:

> *We were waking at 11 pm, 2 am, 5 am, etc to manually [blood glucose] check our daughter. We haven't done that in years. I was having seizures from almost 5 years of not sleeping more than a couple [of] hours at [a] time. Now, we all sleep all night.* [Caregiver of a girl aged 8 years, from the United States; aged 4 years at diagnosis; using Loop]

Even in cases with little improvement in glycemic outcomes, where HbA_{1c} and TIR levels were in or close to the recommended targets before the initiation of open-source AID, caregivers noted that the amount of effort required to achieve these results was significantly diminished. As this point highlights, the data repeatedly pointed to the ways in which physical outcomes are inextricably intertwined with emotional outcomes when considering diabetes management.

Topic 3: Challenges

Difficulties in accessing compatible hardware have frequently been reported. This was mainly associated with differences in the availability of insulin pumps and sensors and reimbursement policies among countries and also with out-of-pocket expenses. Some participants raised concerns regarding access to components and financial aspects of maintaining their open-source AID system in the future:

> *We were concerned about the cost of sensors. They are not covered by private health here and it cost approximately US[D] 5000 a year when we started. Now, kids are covered, but when they turn 21 that ends. We are still worried about covering that bill in the future.* [Caregiver of a boy aged 13 years, from Australia; aged 6 years at diagnosis; using Loop for 2.5 years]

Understanding the documentation and initial setup process is time consuming and challenging, especially for caregivers with little pre-existing knowledge in technology. Ultimately, the complex setup procedure led to a better understanding of the functionalities of open-source AID, enabling caregivers to better respond to technical issues when they occurred. Being part of the *#WeAreNotWaiting* community, caregivers felt gratitude for the available peer support and resources to help with the technical and practical aspects.

Once the setup was successfully managed, the next perceived challenge was the iterative determination of the appropriate settings and therapy parameters to generate satisfactory results. This “fine-tuning” was described as requiring considerable time and endurance. The need to carry around additional devices (eg, a microcontroller or bridge device to remotely communicate between the phone and insulin pump) poses further practical challenges for children in daily life.

The attitudes of HCPs involved in diabetes care of children were described as mixed, ranging from proactive support to refusal:

> *After detailed research, the reserved position of our center could not stop us either. In the past year, I have repeatedly had the impression of knowing more about the disease and the possible forms of therapy than the doctors at our center.* [Caregivers of a girl aged 10 years, from Germany; aged 6 years at diagnosis; using Loop for 1 year]

Despite these reported clinical and quality of life improvements, some expressed uncertainty arising from a lack of support from health care providers. Consequently, a family decided not to disclose the use of open-source AID to their health care team, which caused feelings of isolation, disappointment, and misunderstanding:

> *Although we didn't talk openly about looping, [our diabetologist] has repeatedly emphasized that we only want the DANA RS pump for looping. [...] We[...] now hide the loop, which cannot be a permanent state. We need medical care in which we can communicate openly.* [Caregiver of a boy aged 12 years, from Germany; aged 11 years of age at diagnosis; using AndroidAPS for 2 weeks]

Topic 4: Sources of Support

Participants frequently approached the *#WeAreNotWaiting* community for their support. Social media groups play a key role, where many users share their experiences, respond to questions, discuss related topics, and provide peer support. These were also sources of reassurance in cases of concerns or uncertainties. The extent and quality of peer support available was often a key factor in their decision-making, establishing a sense of trust in the systems, even in the absence of health care provider support or regulatory approval:

> *So in that same Facebook group I started to learn about DIY artificial pa[n]creases and I joined another, international group called Looped to learn more. I then asked around and I was told that OpenAPS was the most advanced of the 3 options and decided to go for that.* [Caregiver of a girl aged 10 years, from Finland; aged 7 years at diagnosis; using OpenAPS for 9 months]

Besides the peer support caregivers found on the web, they attended in-person meetings and local meet-ups with members of the community. Lectures, workshops, and public presentations of open-source AID developers, researchers, and other users and parents enhanced their motivation to start their own journey toward open-source AID. Key individuals who were integral in the development of open-source AID are personally named on a number of occasions. The perceived integrity and altruism of these individuals were also key in creating a sense of confidence and trust in the systems:

> *I found Tim Street's Diabettech website and started following him on Twitter [...]. He was coming to speak at a medical conference in Edinburgh and was going to a [type 1] meet up. I gate-crashed the meet in the pub and had to wait until the end[...] I asked him to show me his pancreas! [...] My son and I flew to London and we got going that evening.* [Caregiver of a boy aged 12 years, from the United Kingdom; aged 8 years at diagnosis; using OpenAPS]

Although HCPs could not prescribe open-source AID systems owing to the absence of regulatory approvals, some were very supportive of the children’s and caregivers’ decision to use open-source AID. Support by HCPs has mostly been reported regarding access to compatible components, such as specific insulin pumps and CGM types that are prescriptible. In a small number of cases, individual caregivers reported that their health care provider initiated a discussion about open-source AID and directed them to relevant sources of information. Conversely,

a lack of support from HCPs was also articulated in a number of accounts, although this usually took the form of "turning a blind eye" and passivity, and very few reported being actively opposed by their health care provider. Where such cases did occur, it usually took the form of preventing caregivers from acquiring the hardware needed to set up an open-source AID system.

Discussion

Principal Findings

In this qualitative analysis, we described the emotional and physical health impact of open-source AID use in children or adolescents and their caregivers, as well as their perceived challenges and sources of support.

Overall, caregivers reported a range of emotions before and after the initiation of open-source AID use. Before initiation, for example, participants reported frustration and dissatisfaction with their existing diabetes management solutions and anticipation and excitement—sometimes marked with anxiety and trepidation—at the prospect of using an open-source AID. Likewise, the experience of using open-source AID evoked both great joy and relief, but this was also tinged, for some, with frustration and worry. Improvements in children's diabetes management, glycemic outcomes, physical health beyond diabetes, and emotional well-being were highlighted in the narratives. Furthermore, sleep quality and quality of life improved for both children and caregivers. The initial challenges were difficulties in accessing the required components, lack of confidence in technical skills for setup and maintenance, concerns about the response from health care teams, and the wider social environment of the family. Later, finding and "fine-tuning" of the right therapy settings, as well as the impracticality of carrying additional devices for the children, were described. The *#WeAreNotWaiting* web-based community was frequently approached as the primary source of information as well as emotional and practical support.

This study can inform stakeholders regarding the unmet needs of children and adolescents with T1D regarding the therapeutic options available to them. Furthermore, our findings highlight how children might benefit from customizable open-source AID systems where commercial options are not accessible, approved for certain age groups, or limited in their functionality to cover the lower and variable insulin requirements of children.

Comparison With Prior Work

The ethical and legal aspects of the off-label use of unregulated medical devices in children and adolescents are multifaceted and complex. Although the off-label use of pharmaceuticals is both common practice and a necessity in pediatrics, it is still uncommon in medical devices. HCPs were sometimes perceived to be reticent in their support of the decision to use open-source AID. This reticence is understandable given that many HCPs, as indicated by the caregivers in this study, had very little knowledge of the systems and uncertainty regarding what legal ramifications there might be in providing support for a system not approved by regulatory bodies. Following a number of position papers from several local diabetes organizations, a group of international HCPs recently provided an international consensus statement for practical guidance on the safe and ethical use of open-source in clinical settings [18]. The consensus encourages colleagues to learn about all treatment options that could help people with diabetes, including open-source AID, and to support individual decisions to use open-source AID for treatment, as long as benefits and risks are understood. In addition, children's welfare must always be considered by caregivers and HCPs, with their assent and engagement [18].

Although there are numerous studies about the clinical outcomes of the use of open-source and commercially available AID systems in adults and children [13-15,22-24], there is yet very limited knowledge about the lived experiences and psychological antecedents or consequences leading to the use of and with AID. To the best of our knowledge, this is the most extensive study on lived experiences in children and adolescents using open-source AID, and their caregivers and families, conducted at a multinational level. Our findings are in line with other studies that indicated a reduced burden of diabetes in users of commercial and open-source AID [8,25-31]. Caregivers' sleep and mental and physical health in the context of their children's diabetes remains an underresearched area. A reduced burden on caregivers of young children was previously identified as the main outcome of the use of a commercial AID system [8]. The DIWHY survey was conducted between 2018 and 2019. At the time, only one commercially developed AID system was approved and made available in the United States. We did not explicitly ask for this information, although with only 4 participants from the United States, it can be assumed that most of the participants did not have access to commercial AID. Furthermore, open-source AID systems have continuously improved over time with respect to usability and device interoperability. For example, the need to carry around additional hardware may no longer be applied in the recent versions of AndroidAPS, FreeAPS, and Loop. We suggest further research in this field to provide a better understanding of the full psychosocial and economic impact of any kind of AID, as well as the challenges in the access and use of these systems.

Strengths and Limitations

This study has several strengths and limitations. Of particular strength is its patient and public involvement in the study design process and its multinational scope. Limitations include that the anonymous study design did not allow participants to follow up for clarification or further questions, to strictly follow the General Data Protection Regulation guidelines. A selection bias may be present with the survey only being available in German and English, which may have excluded users not proficient in these languages in the first place. Furthermore, those within the sample might not have responded in detail, or not at all, to the open-ended questions owing to language barriers among other factors. In addition, a significant proportion of the OPEN team was German, with strong links to the German diabetes community; therefore, the teams' ability to reach people was particularly high in that country. Finally, the challenges in building and setting up an open-source AID had to be overcome by caregivers with perseverance and self-motivation in the first

place, potentially resulting in a selected population that limits broad generalizations to all people with diabetes.

Conclusions

With frequent changes in insulin requirements, glycemic variability due to counterregulatory hormones, and physical activity, children are ideal candidates for AID. Although the uptake of insulin pumps and CGM is high among children in countries where access to diabetes technology is facilitated, the uptake of AID in children is protracted owing to the license status of commercially available AID systems. However, their efficacy in young children and those with low insulin requirements remains limited. Furthermore, glycemic outcome improvements in the off-label use of commercial AID systems by very young children are suboptimal, although they experience similar glycemic improvements as older children, adolescents, and adults with commercial systems approved for their age [7] and with open-source AID [13-16]. Our findings indicate a transformative impact of open-source AID in children and adolescents of various ages on their emotional and physical health, as well as their and their caregivers' sleep and quality of life. They further highlight how remote monitoring and control are perceived by parents to be safe and how the children are provided with greater autonomy.

Similar to commercial AID systems, there remains much room for improvement in open-source AID systems, and further research is needed to improve the effectiveness of algorithms and usability of AID systems in general, particularly in young children where approved therapy options remain limited. To achieve this, concerted efforts are required using a multi-stakeholder approach, an approach in which the diverse and valuable experiences of caregivers and children who have opted to move into the vanguard of AID need to be heard and appreciated.

Acknowledgments

The authors would like to thank the members of the *#WeAreNotWaiting* community and their families for their collaboration and support in cocreating and disseminating this study.

The authors acknowledge the support of Dana Lewis, developer of OpenAPS, and Adrian Tappe, developer of AndroidAPS, for the study design. We further acknowledge the support of Dr Henriette Langstrup as mentor for the interpretation and discussion of the findings of this study.

The authors acknowledge support from the German Research Foundation and the Open Access Publication Fund of Charité—Universitätsmedizin Berlin.

Conflicts of Interest

KB received fees for medical consulting and public speaking from Roche Diabetes Care, Dexcom, Medtronic Diabetes, Diabeloop, Novo Nordisk, Sanofi Diabetes, Diabetes Center Berne, Abbott and BCG Digital Ventures; outside the submitted work. KR received fees for medical consulting and public speaking from Dexcom, Abbott, Lilly Diabetes Care, Novo Nordisk (Germany), and Springer Healthcare United Kingdom; outside the submitted work.
KR received fees for medical consulting and public speaking from Dexcom, Abbott, Lilly Diabetes Care, Novo Nordisk (Germany), and Springer Healthcare United Kingdom; outside the submitted work.

Multimedia Appendix 1
Questionnaires for caregivers of children and adolescents with diabetes, using open-source automated insulin delivery.
[PDF File (Adobe PDF File), 92 KB - jmir_v24i7e37120_app1.pdf]

Multimedia Appendix 2
Demographic and clinical characteristics of participants of the DIWHY study who have not responded to any of the open-ended questions.
[DOCX File , 16 KB - jmir_v24i7e37120_app2.docx]

References

1. DiMeglio LA, Acerini CL, Codner E, Craig ME, Hofer SE, Pillay K, et al. ISPAD Clinical Practice Consensus Guidelines 2018: glycemic control targets and glucose monitoring for children, adolescents, and young adults with diabetes. Pediatr Diabetes 2018 Oct 01;19:105-114. [doi: 10.1111/pedi.12737]
2. Szadkowska A, Pietrzak I, Mianowska B, Bodalska-Lipińska J, Keenan HA, Toporowska-Kowalska E, et al. Insulin sensitivity in type 1 diabetic children and adolescents. Diabet Med 2008 Mar;25(3):282-288. [doi: 10.1111/j.1464-5491.2007.02357.x] [Medline: 18279410]
3. Piona C, Marigliano M, Mozzillo E, Franzese A, Zanfardino A, Iafusco D, et al. Long-term glycemic control and glucose variability assessed with continuous glucose monitoring in a pediatric population with type 1 diabetes: determination of optimal sampling duration. Pediatr Diabetes 2020 Dec 28;21(8):1485-1492. [doi: 10.1111/pedi.13115] [Medline: 32935887]

4. Karges B, Rosenbauer J, Holterhus PM, Beyer P, Seithe H, Vogel C, et al. Hospital admission for diabetic ketoacidosis or severe hypoglycemia in 31,330 young patients with type 1 diabetes. Eur J Endocrinol 2015 Sep;173(3):341-350. [doi: 10.1530/EJE-15-0129] [Medline: 26088822]
5. Foster NC, Beck RW, Miller KM, Clements MA, Rickels MR, DiMeglio LA, et al. State of type 1 diabetes management and outcomes from the T1D exchange in 2016-2018. Diabetes Technol Ther 2019 Feb;21(2):66-72 [FREE Full text] [doi: 10.1089/dia.2018.0384] [Medline: 30657336]
6. Bronner MB, Peeters MA, Sattoe JN, van Staa A. The impact of type 1 diabetes on young adults' health-related quality of life. Health Qual Life Outcomes 2020 May 12;18(1):137 [FREE Full text] [doi: 10.1186/s12955-020-01370-8] [Medline: 32398086]
7. Tauschmann M, Allen J, Nagl K, Fritsch M, Yong J, Metcalfe E, KidsAP Consortium. Home use of day-and-night hybrid closed-loop insulin delivery in very young children: a multicenter, 3-week, randomized trial. Diabetes Care 2019 Apr;42(4):594-600. [doi: 10.2337/dc18-1881] [Medline: 30692242]
8. Musolino G, Dovc K, Boughton CK, Tauschmann M, Allen JM, Nagl K, Kidsap Consortium. Reduced burden of diabetes and improved quality of life: experiences from unrestricted day-and-night hybrid closed-loop use in very young children with type 1 diabetes. Pediatr Diabetes 2019 Sep 13;20(6):794-799 [FREE Full text] [doi: 10.1111/pedi.12872] [Medline: 31140654]
9. Forlenza GP, Buckingham BA, Brown SA, Bode BW, Levy CJ, Criego AB, et al. First outpatient evaluation of a tubeless automated insulin delivery system with customizable glucose targets in children and adults with type 1 diabetes. Diabetes Technol Ther 2021 Jun 01;23(6):410-424 [FREE Full text] [doi: 10.1089/dia.2020.0546] [Medline: 33325779]
10. Schoelwer MJ, Robic JL, Gautier T, Fabris C, Carr K, Clancy-Oliveri M, et al. Safety and efficacy of initializing the control-IQ artificial pancreas system based on total daily insulin in adolescents with type 1 diabetes. Diabetes Technol Ther 2020 Aug 01;22(8):594-601 [FREE Full text] [doi: 10.1089/dia.2019.0471] [Medline: 32119790]
11. Varimo T, Pulkkinen M, Hakonen E, Hero M, Miettinen PJ, Tuomaala A. First year on commercial hybrid closed-loop system-experience on 111 children and adolescents with type 1 diabetes. Pediatr Diabetes 2021 Sep 28;22(6):909-915 [FREE Full text] [doi: 10.1111/pedi.13235] [Medline: 34015178]
12. Salehi P, Roberts AJ, Kim GJ. Efficacy and safety of real-life usage of MiniMed 670g Automode in children with type 1 diabetes less than 7 years old. Diabetes Technol Ther 2019 Aug;21(8):448-451. [doi: 10.1089/dia.2019.0123] [Medline: 31166801]
13. Lum JW, Bailey RJ, Barnes-Lomen V, Naranjo D, Hood KK, Lal RA, et al. A real-world prospective study of the safety and effectiveness of the loop open source automated insulin delivery system. Diabetes Technol Ther 2021 May 01;23(5):367-375 [FREE Full text] [doi: 10.1089/dia.2020.0535] [Medline: 33226840]
14. Braune K, Gajewska KA, Thieffry A, Lewis DM, Froment T, O'Donnell S, et al. Why #WeAreNotWaiting-motivations and self-reported outcomes among users of open-source automated insulin delivery systems: multinational survey. J Med Internet Res 2021 Jun 07;23(6):e25409 [FREE Full text] [doi: 10.2196/25409] [Medline: 34096874]
15. Braune K, O'Donnell S, Cleal B, Lewis D, Tappe A, Willaing I, et al. Real-world use of do-it-yourself artificial pancreas systems in children and adolescents with type 1 diabetes: online survey and analysis of self-reported clinical outcomes. JMIR Mhealth Uhealth 2019 Jul 30;7(7):e14087 [FREE Full text] [doi: 10.2196/14087] [Medline: 31364599]
16. Petruzelkova L, Jiranova P, Soupal J, Kozak M, Plachy L, Neuman V, et al. Pre-school and school-aged children benefit from the switch from a sensor-augmented pump to an AndroidAPS hybrid closed loop: a retrospective analysis. Pediatr Diabetes 2021 Jun 25;22(4):594-604 [FREE Full text] [doi: 10.1111/pedi.13190] [Medline: 33576551]
17. O'Donnell S, Lewis D, Marchante Fernández M, Wäldchen M, Cleal B, Skinner T, et al. Evidence on user-led innovation in diabetes technology (The OPEN Project): protocol for a mixed methods study. JMIR Res Protoc 2019 Nov 19;8(11):e15368. [doi: 10.2196/15368]
18. Braune K, Lal RA, Petruželková L, Scheiner G, Winterdijk P, Schmidt S, et al. Open-source automated insulin delivery: international consensus statement and practical guidance for health-care professionals. Lancet Diabetes Endocrinol 2022 Jan;10(1):58-74. [doi: 10.1016/s2213-8587(21)00267-9]
19. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
20. Brooks J, McCluskey S, Turley E, King N. The utility of template analysis in qualitative psychology research. Qual Res Psychol 2015 Apr 03;12(2):202-222 [FREE Full text] [doi: 10.1080/14780887.2014.955224] [Medline: 27499705]
22. Gawrecki A, Zozulinska-Ziolkiewicz D, Michalak MA, Adamska A, Michalak M, Frackowiak U, et al. Safety and glycemic outcomes of do-it-yourself AndroidAPS hybrid closed-loop system in adults with type 1 diabetes. PLoS One 2021 Apr 5;16(4):e0248965 [FREE Full text] [doi: 10.1371/journal.pone.0248965] [Medline: 33819289]
23. Melmer A, Züger T, Lewis DM, Leibrand S, Stettler C, Laimer M. Glycaemic control in individuals with type 1 diabetes using an open source artificial pancreas system (OpenAPS). Diabetes Obes Metab 2019 Oct 05;21(10):2333-2337. [doi: 10.1111/dom.13810] [Medline: 31183929]
24. Jeyaventhan R, Gallen G, Choudhary P, Hussain S. A real-world study of user characteristics, safety and efficacy of open-source closed-loop systems and Medtronic 670G. Diabetes Obes Metab 2021 Aug 09;23(8):1989-1994. [doi: 10.1111/dom.14439] [Medline: 33999488]

25. Schipp J, Skinner T, Holloway E, Scibilia R, Langstrup H, Speight J, et al. How adults with type 1 diabetes are navigating the challenges of open-source artificial pancreas systems: a qualitative study. Diabetes Technol Ther 2021 Aug 01;23(8):546-554 [FREE Full text] [doi: 10.1089/dia.2020.0652] [Medline: 33720767]
26. Marshall DC, Holloway M, Korer M, Woodman J, Brackenridge A, Hussain S. Do-it-yourself artificial pancreas systems in type 1 diabetes: perspectives of two adult users, a caregiver and three physicians. Diabetes Ther 2019 Oct 22;10(5):1553-1564 [FREE Full text] [doi: 10.1007/s13300-019-00679-y] [Medline: 31440989]
27. Hendrieckx C, Poole LA, Sharifi A, Jayawardene D, Loh MM, Horsburgh JC, et al. "It is definitely a game changer": a qualitative study of experiences with in-home overnight closed-loop technology among adults with type 1 diabetes. Diabetes Technol Ther 2017 Jul;19(7):410-416. [doi: 10.1089/dia.2017.0007] [Medline: 28537437]
28. Barnard KD, Wysocki T, Allen JM, Elleri D, Thabit H, Leelarathna L, et al. Closing the loop overnight at home setting: psychosocial impact for adolescents with type 1 diabetes and their parents. BMJ Open Diabetes Res Care 2014 Apr 16;2(1):e000025 [FREE Full text] [doi: 10.1136/bmjdrc-2014-000025] [Medline: 25452866]
29. Hng T, Burren D. Appearance of Do-It-Yourself closed-loop systems to manage type 1 diabetes. Intern Med J 2018 Nov;48(11):1400-1404. [doi: 10.1111/imj.14105] [Medline: 30387315]
30. Bisio A, Brown SA, McFadden R, Pajewski M, Yu PL, DeBoer M, et al. Sleep and diabetes-specific psycho-behavioral outcomes of a new automated insulin delivery system in young children with type 1 diabetes and their parents. Pediatr Diabetes 2021 May 08;22(3):495-502. [doi: 10.1111/pedi.13164] [Medline: 33289242]
31. Lawton J, Blackburn M, Rankin D, Allen JM, Campbell FM, Leelarathna L, et al. Participants' experiences of, and views about, daytime use of a day-and-night hybrid closed-loop system in real life settings: longitudinal qualitative study. Diabetes Technol Ther 2019 Mar;21(3):119-127 [FREE Full text] [doi: 10.1089/dia.2018.0306] [Medline: 30720338]

Abbreviations

AID: automated insulin delivery
CGM: continuous glucose monitor
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
HbA1c: hemoglobin A1c
HCP: health care professional
OPEN: Outcomes of Patients' Evidence with Novel, Do-it-Yourself Artificial Pancreas Technology
REDCap: Research Electronic Data Capture
T1D: type 1 diabetes
TIR: time in range

Edited by R Kukafka, G Eysenbach; submitted 07.02.22; peer-reviewed by S Hussain, R Lal, C Campos-Castillo, L Laestadius; comments to author 29.03.22; revised version received 21.04.22; accepted 29.05.22; published 14.07.22.

Please cite as:
Braune K, Krug N, Knoll C, Ballhausen H, Thieffry A, Chen Y, O'Donnell S, Raile K, Cleal B
Emotional and Physical Health Impact in Children and Adolescents and Their Caregivers Using Open-source Automated Insulin Delivery: Qualitative Analysis of Lived Experiences
J Med Internet Res 2022;24(7):e37120
URL: https://www.jmir.org/2022/7/e37120
doi:10.2196/37120
PMID:35834298

Original Paper

Prediction of Maternal Hemorrhage Using Machine Learning: Retrospective Cohort Study

Jill M Westcott[1], MD, MSc; Francine Hughes[2], MD; Wenke Liu[3,4], PhD; Mark Grivainis[3,4], MSc; Iffath Hoskins[1], MD; David Fenyo[3,4], PhD

[1]Division of Maternal-Fetal Medicine, Department of Obstetrics and Gynecology, New York University Langone Health, New York, NY, United States

[2]Montefiore Medical Center, Albert Einstein College of Medicine, Bronx, NY, United States

[3]Institute for Systems Genetics, New York University Grossman School of Medicine, New York, NY, United States

[4]Department of Biochemistry and Molecular Pharmacology, New York University Grossman School of Medicine, New York, NY, United States

Corresponding Author:
Jill M Westcott, MD, MSc
Division of Maternal-Fetal Medicine, Department of Obstetrics and Gynecology
New York University Langone Health
462 First Avenue
New York, NY, 10016
United States
Phone: 1 2122637000
Email: jmwestcott@cmh.edu

Abstract

Background: Postpartum hemorrhage remains one of the largest causes of maternal morbidity and mortality in the United States.

Objective: The aim of this paper is to use machine learning techniques to identify patients at risk for postpartum hemorrhage at obstetric delivery.

Methods: Women aged 18 to 55 years delivering at a major academic center from July 2013 to October 2018 were included for analysis (N=30,867). A total of 497 variables were collected from the electronic medical record including the following: demographic information; obstetric, medical, surgical, and family history; vital signs; laboratory results; labor medication exposures; and delivery outcomes. Postpartum hemorrhage was defined as a blood loss of ≥1000 mL at the time of delivery, regardless of delivery method, with 2179 (7.1%) positive cases observed. Supervised learning with regression-, tree-, and kernel-based machine learning methods was used to create classification models based upon training (21,606/30,867, 70%) and validation (4630/30,867, 15%) cohorts. Models were tuned using feature selection algorithms and domain knowledge. An independent test cohort (4631/30,867, 15%) determined final performance by assessing for accuracy, area under the receiver operating curve (AUROC), and sensitivity for proper classification of postpartum hemorrhage. Separate models were created using all collected data versus models limited to data available prior to the second stage of labor or at the time of decision to proceed with cesarean delivery. Additional models examined patients by mode of delivery.

Results: Gradient boosted decision trees achieved the best discrimination in the overall model. The model including all data mildly outperformed the second stage model (AUROC 0.979, 95% CI 0.971-0.986 vs AUROC 0.955, 95% CI 0.939-0.970). Optimal model accuracy was 98.1% with a sensitivity of 0.763 for positive prediction of postpartum hemorrhage. The second stage model achieved an accuracy of 98.0% with a sensitivity of 0.737. Other selected algorithms returned models that performed with decreased discrimination. Models stratified by mode of delivery achieved good to excellent discrimination but lacked the sensitivity necessary for clinical applicability.

Conclusions: Machine learning methods can be used to identify women at risk for postpartum hemorrhage who may benefit from individualized preventative measures. Models limited to data available prior to delivery perform nearly as well as those with more complete data sets, supporting their potential utility in the clinical setting. Further work is necessary to create successful models based upon mode of delivery and to validate the findings of this study. An unbiased approach to hemorrhage risk prediction may be superior to human risk assessment and represents an area for future research.

(J Med Internet Res 2022;24(7):e34108) doi:10.2196/34108

KEYWORDS

predictive modeling; maternal morbidity; postpartum hemorrhage; machine learning; obstetrics; pregnancy; post partum; maternal

Introduction

Postpartum hemorrhage is the leading cause of maternal mortality worldwide [1]. In the United States, the rate of postpartum hemorrhage continues to rise, complicating nearly 3% of deliveries [2]. Mothers with severe hemorrhage may require blood transfusion, hysterectomy, or intensive care unit admission with a select number of cases proving fatal. Postpartum hemorrhage that leads to blood transfusion is the leading cause of severe maternal morbidity in the United States [3]. Stewardship of blood resources and minimizing hemorrhage-related morbidity remain ongoing efforts as blood transfusion is not without risk. By predicting patients at risk for significant blood loss, prophylactic measures may be instituted to avoid maternal morbidity and mortality.

A number of risk factors for postpartum hemorrhage have been established, including previous postpartum hemorrhage, multifetal gestation, pre-eclampsia, augmented labor, fetal macrosomia, operative vaginal delivery, and complex lacerations, as well as other factors [4]. Previous models for prediction of postpartum hemorrhage have been developed [5-7], but validation of these among different populations and at different time points within the labor process has been limited. A machine learning study using administrative data provided poor discrimination for predicting the need for hospital readmission due to postpartum hemorrhage in the first 12 weeks postpartum [8]. Prediction of postpartum hemorrhage remains a challenge for the obstetric provider, and further work is necessary using modern modeling methods.

The field of machine learning has recently seen a rapid development of methods that support unbiased learning from data. Supervised learning involves processing information to predict from examples with a known outcome, often for the purpose of estimating risk in examples where the outcome is not known [9]. Multiple applications for machine learning exist within medicine; however, to date, they have not been widely used in the field of obstetrics. By using the power of modern predictive modeling for postpartum hemorrhage, we aim to better identify those patients at increased risk for obstetric hemorrhage to avoid maternal morbidity and mortality. Identifying patients at the highest risk of postpartum hemorrhage will enable providers to reduce the cost and morbidity associated with postpartum hemorrhage and ultimately improve patient outcomes.

Methods

Ethics Approval

Institutional Review Board approval was obtained from New York University Langone Health (approval number s18-01798).

Study Population

This was a retrospective cohort study conducted at a single tertiary care center. Women aged 18 to 55 years delivering at New York University Langone Health Tisch Hospital from July 1, 2013, to October 31, 2018, were included for analysis. Patients not meeting age parameters as well as those cases in which a blood loss value was either not available or not recorded were excluded. All patients not meeting exclusion criteria were included in the study.

Study Design and Model Development

A total of 497 variables were collected from unique sources within the electronic medical record including the following: demographic information; obstetric, medical, surgical, and family history; vital signs; laboratory results; labor exposures; and delivery outcomes (Multimedia Appendix 1). Postpartum hemorrhage was defined as a blood loss of ≥1000 mL at the time of delivery, as recommended by the American College of Obstetricians and Gynecologists revitalize program [10].

The delivery cohort was randomly split into training (21,606/30,867, 70% of total cohort) and validation (4630/30,867, 15%) sets for model creation. Using the R software (version 3.5.1; R Foundation for Statistical Computing), supervised learning with regression-, tree-, and kernel-based machine learning methods was used to create classification models, using each method for every model assessed. The models were tuned using recursive feature selection, selection by filtering, observing feature importance, and domain knowledge. The model parameters were customized and examined to produce optimal results. An independent test cohort (4631/30,867, 15%) determined the final performance by assessing for accuracy, area under the receiver operating curve (AUROC), and sensitivity for the proper classification of postpartum hemorrhage.

The initial model included variables that contained information that would be feasible to obtain prior to delivery (ie, relevant historical information, objective data present within the inpatient and outpatient chart, and diagnoses associated with the patient's delivery encounter entered within 24 hours following delivery). A secondary model was created limited to data strictly expected to be available prior to the second stage of labor or at the time of decision to proceed with cesarean delivery, as this was likely the more clinically useful tool. Additional models were created for patients undergoing cesarean and vaginal delivery.

The selection of appropriate variables for inclusion was made by a single obstetric provider with experience and knowledge of the electronic medical record. The number of initial variables in each model differed according to clinical applicability. The variables were processed according to the provider's assessment of the clinical scenario noted for each patient.

Results

A total of 30,867 patients met the inclusion criteria, and 2179 (7.1%) cases met the criteria for postpartum hemorrhage. Patient characteristics are detailed in Table 1. Cesarean delivery was noted in 27.6% (n=8534) of the patients, and unknown mode of delivery in 0.1% (n=19), with the remainder (n=22,314, 72.3%) undergoing vaginal delivery. The rate of postpartum

hemorrhage by mode of delivery was 20.8% (1776/8534) for cesarean delivery, 21.1% (4/19) for unknown mode of delivery, and 1.8% (399/22,314) for vaginal delivery. The average gestational age was 274.6 (range 107-303) days, and the average patient age was 32.7 (range 18-55) years. The delivery cohort was split into training, validation, and test cohorts for initial model creation containing 21,606 (70%), 4630 (15%), and 4631 (15%) patients, respectively.

Table 1. Patient cohort characteristics.

Variable	Value	
	Overall (N=30,867)	PPH[a] (n=2179)
Age (years), mean (range, SD)	32.7 (18-55, 5.3)	34.3 (18-54, 5.4)
Gestation (days)	274.6	272.2
BMI (kg/m^2)	28.5	30.0
Vaginal delivery, n (%)	22,314 (72.3)	399 (18.3)
Cesarean delivery, n (%)	8534 (27.6)	1776 (81.5)
Unknown mode of delivery, n (%)	19 (0.1)	4 (0.2)
Operative vaginal delivery, n (%)	1502 (4.9)	51 (2.3)
Vaginal birth after cesarean, n (%)	1019 (3.3)	24 (1.1)
Multiple gestation, n (%)	143 (0.5)	35 (1.6)
Current tobacco use, n (%)	142 (0.5)	10 (0.5)
Maternal diabetes, n (%)	1544 (5.0)	172 (7.9)
Maternal pre-eclampsia or eclampsia, n (%)	508 (1.6)	65 (3.0)
IUFD[b], n (%)	111 (0.4)	11 (0.5)
Labor induced, n (%)	16,962 (55.0)	927 (42.5)
Primiparous, n (%)	16,858 (54.6)	1404 (64.4)
Primary cesarean delivery, n (%)	7070 (82.8)	1564 (88.1)
Cesarean performed prior to labor or rupture, n (%)	4801 (56.3)	880 (49.5)

[a]PPH: postpartum hemorrhage.

[b]IUFD: intrauterine fetal demise.

The initial model included a total of 280 variables. Logistic regression, random forest, gradient boosted decision trees), and support vector machine models were generated to create a representative sample of different methods. Gradient boosted decision trees achieved the best discrimination among the initial models, performing with an AUROC of 0.979 (95% CI 0.971-0.986) and an accuracy of 98.1%. Sensitivity for this model was 0.763 (95% CI 0.712-0.809, Table 2). Other models performed less successfully. The optimal model included 212 features (Multimedia Appendix 2).

Table 2. Optimal performance of all models using gradient boosted decision trees.

Model	Accuracy	AUROC[a]	Sensitivity
Initial	0.981	0.979	0.763
Second stage	0.980	0.955	0.737
Cesarean delivery	0.818	0.737	0.320
Vaginal delivery (1 hour[b])	0.982	0.837	0.254
Vaginal delivery (second stage)	0.983	0.846	0.254

[a]AUROC: area under the receiver operating curve.

[b]Within 1 hour of presumed admission for vaginal delivery.

The data set was then trimmed to include only those variables (123 in total) available prior to the second stage of labor or at the time of decision to proceed with cesarean delivery. A similar representative sample of modeling methods was used, with gradient boosted decision trees again achieving the best discrimination, noting an AUROC of 0.955 (95% CI 0.939-0.970) and an accuracy of 98.0%. Sensitivity for this model was 0.737 (95% CI 0.684-0.785; Table 2). This model included a total of 28 features (Textbox 1). The most important features included body mass index, admission hematocrit,

cesarean delivery prior to labor or rupture, scheduling status of cesarean delivery, and admission platelet count.

Textbox 1. Features included in the optimal second stage model.

Laboratory components

- Eosinophils (absolute)
- Hematocrit
- Hemoglobin
- Lymphocytes (absolute)
- Lymphocytes (percent)
- Mean corpuscular hemoglobin
- Mean corpuscular hemoglobin concentration
- Mean corpuscular volume
- Mean platelet volume
- Monocytes (percent)
- Neutrophils (absolute)
- Neutrophils (percent)
- Platelet count
- Red cell distribution width (standard deviation)
- Red blood cell count
- White blood cell count

Nonlaboratory components

- Patient age
- Gestational age
- Systolic blood pressure
- Diastolic blood pressure
- BMI
- Pulse oximetry
- Temperature
- Live birth count
- Baseline fetal heart rate
- Amniotic fluid color
- Cesarean delivery prior to labor or rupture
- Cesarean delivery scheduling status

Additional models focused on classifying patients by mode of delivery. A model examining those patients ultimately delivered by cesarean delivery (n=8534) was created using information available at the time of decision to proceed with cesarean delivery (ie, within 1 hour after presentation for scheduled procedure or following attempted vaginal delivery prior to proceeding to the operating room). A total of 173 variables were included. Using gradient boosted decision trees, the highest performing model contained 76 variables (Multimedia Appendix 3) and achieved an AUROC of 0.737 (95% CI 0.703-0.772). Accuracy of 81.8% and a lower sensitivity of 0.320 were noted (95% CI 0.266-0.377; Table 2).

Two models using data of those patients who underwent vaginal delivery (n=22,333) were also created. The first focused on information available within 1 hour of admission for presumed vaginal delivery and examined 127 variables. The model achieved excellent discrimination, noting an AUROC of 0.837 (95% CI 0.782-0.893) and accuracy of 98.2%; however, sensitivity remained low at 0.254 (95% CI 0.155-0.375). The optimal model included 7 variables (Multimedia Appendix 4). The second model used information available prior to the second stage of labor; thus, it comprised variables related to the patient's labor course, including medication exposure and induction method, if applicable. A total of 176 features were included, resulting in an optimal model with an AUROC of 0.846 (95% CI 0.790-0.902) and accuracy of 98.3%. Sensitivity was again

0.254 (95% CI 0.153-0.379), and the final model included 92 features (Table 2; Multimedia Appendix 5). Both of these models were achieved using gradient boosted decision trees. Third trimester and admission hemoglobin or hematocrit were among the most important features in both vaginal delivery models.

Discussion

Principal Findings

Our study successfully produced a model for predicting postpartum hemorrhage in patients undergoing obstetric delivery. When using only the data available prior to the second stage of labor or at the time to proceed with cesarean delivery, we achieved nearly equal discrimination and sensitivity compared to our more robust initial model, successfully predicting nearly 3 out of every 4 patients who had a postpartum hemorrhage.

Many previously identified risk factors for postpartum hemorrhage were not included in the final model, including multiple gestation, operative vaginal delivery, and history of postpartum hemorrhage, among others. This indicates that many of these factors may not be as contributory to postpartum hemorrhage risk as previously believed, but further work is necessary.

Prior Results

Postpartum hemorrhage is a known cause of significant maternal morbidity and mortality in the United States and remains difficult to predict. Few existing studies have used machine learning methods to identify patients at risk for postpartum hemorrhage with minimal success [5-8]. A recently published model used a large cohort from the US Consortium for Safe Labor and achieved excellent discrimination, although its utility in the clinical setting is limited given its retrospective nature without prospective validation [11]. This study used 55 predictor variables, indicating a less robust data set than what was curated for our model. Our study represents the largest cohort to date to generate a predictive risk model using data directly abstracted from the electronic medical record that is applicable in a targeted population.

When stratified by delivery method, our models noted a decreased sensitivity. While this may appear in contradiction to the expected results, it is understandable because the majority of postpartum hemorrhages occurred in those patients who underwent cesarean delivery. This is further reflected by examining the most important features in our final second-stage model.

Clinical Implications

The ability to predict patients at risk for postpartum hemorrhage using readily available information represents an area of substantial clinical opportunity. Integrating a model such as ours into clinical practice will give providers the real-time capability to assess a patient's risk of hemorrhage. Targeted intervention, such as prophylactic administration of uterotonic medication, availability of blood products, and even potentially transferring patients to a center offering a higher level of maternal care [12] is a consideration for those deemed at risk.

Strengths and Limitations

The strengths of this study include the use of modern supervised machine learning techniques in a clinical condition that has not been extensively explored with this approach. This data set represents the largest directly derived cohort to use these techniques. Additionally, the inclusion of nearly 500 variables in the data set provides a robust cohort from which to create the model, and this size has not been previously seen in the literature. As machine learning methods are centered upon improving performance with increasing inputs, this lends to a superior model. Since having a large number of overfitting variables is a concern, this must be considered when determining the optimal model. A slight decrease in accuracy may be necessary to select a model with less concern for overfitting. The use of independent validation and test cohorts also supports the strength and lack of bias in our model.

Limitations include the retrospective nature of this study as well as the use of a population from a single tertiary center. Given regional variations in patient populations, our results may not be generalizable to the US population at large, and we do note a higher rate of postpartum hemorrhage in our cohort than previously described. It is unclear why the rate was higher in our population, but it may be partially explained by the referral nature of our tertiary center, leading to care of a larger number of patients at high risk at baseline. Further validation with an outside cohort and prospective validation among our patient population is necessary.

The use of the electronic medical record is an additional limitation to our study. Differences or duplications in both location and format of inputs have the potential to impair the accuracy of our abstracted data. We are unable to assess the performance or bias of this model across race as this is a variable inputted by the clinical staff; thus, we are unable to validate its accuracy. The variables related to diagnosis codes are entirely dependent upon provider input, and all applicable conditions may not have been entered. However, with a large data set, machine learning algorithms should be able to overcome this deficit as features with a high level of contribution to the outcome should persist when feature selection is implemented.

The class imbalance of positive or negative cases for postpartum hemorrhage in the data set is inevitable given the relatively low incidence of this condition in clinical practice. This was particularly evident in the support vector machine models where every patient was predicted not to be at risk for postpartum hemorrhage. The use of a weighted loss could be considered to compensate for this imbalance.

Conclusions

In conclusion, machine learning methods are a less used approach in obstetrics and can be used to identify women at risk for postpartum hemorrhage who may benefit from individualized preventative measures. Models limited to data available prior to delivery perform nearly as well as those with more complete data sets, identifying nearly three-quarters of

patients at risk, supporting their potential utility in the clinical setting.

Acknowledgments

This research was funded through an NYU CTSA grant UL1 TR001445 from the National Center for Advancing Translational Sciences, National Institutes of Health. The authors wish to thank the members of NYU Langone Health DataCore for their assistance in this project.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary material 1: 497 variables abstracted.
[DOCX File , 25 KB - jmir_v24i7e34108_app1.docx]

Multimedia Appendix 2
Supplementary material 2: 212 variables abstracted for initial model (*** = top 5 importance).
[DOCX File , 20 KB - jmir_v24i7e34108_app2.docx]

Multimedia Appendix 3
Supplementary material 3: 76 variables abstracted for cesarean delivery model (*** = top 5 importance).
[DOCX File , 15 KB - jmir_v24i7e34108_app3.docx]

Multimedia Appendix 4
Supplementary material 4: 7 variables abstracted for 1 hour vaginal delivery model (*** = top 5 importance).
[DOCX File , 13 KB - jmir_v24i7e34108_app4.docx]

Multimedia Appendix 5
Supplementary material 5: 92 variables abstracted for 2nd stage vaginal delivery model (*** = top 5 importance).
[DOCX File , 16 KB - jmir_v24i7e34108_app5.docx]

References

1. Say L, Chou D, Gemmill A, Tunçalp Ö, Moller A, Daniels J, et al. Global causes of maternal death: a WHO systematic analysis. Lancet Glob Health 2014 Jun;2(6):e323-e333 [FREE Full text] [doi: 10.1016/S2214-109X(14)70227-X] [Medline: 25103301]
2. Callaghan WM, Kuklina EV, Berg CJ. Trends in postpartum hemorrhage: United States, 1994-2006. Am J Obstet Gynecol 2010 Apr;202(4):353.e1-353.e6. [doi: 10.1016/j.ajog.2010.01.011] [Medline: 20350642]
3. Creanga AA, Berg CJ, Ko JY, Farr SL, Tong VT, Bruce FC, et al. Maternal mortality and morbidity in the United States: where are we now? J Womens Health (Larchmt) 2014 Jan;23(1):3-9 [FREE Full text] [doi: 10.1089/jwh.2013.4617] [Medline: 24383493]
4. Oyelese Y, Ananth CV. Postpartum hemorrhage: epidemiology, risk factors, and causes. Clin Obstet Gynecol 2010 Mar;53(1):147-156. [doi: 10.1097/GRF.0b013e3181cc406d] [Medline: 20142652]
5. Rubio-Álvarez A, Molina-Alarcón M, Arias-Arias Á, Hernández-Martínez A. Development and validation of a predictive model for excessive postpartum blood loss: A retrospective, cohort study. Int J Nurs Stud 2018 Mar;79:114-121. [doi: 10.1016/j.ijnurstu.2017.11.009] [Medline: 29223625]
6. Helman S, Drukker L, Fruchtman H, Ioscovich A, Farkash R, Avitan T, et al. Revisit of risk factors for major obstetric hemorrhage: insights from a large medical center. Arch Gynecol Obstet 2015 Oct;292(4):819-828. [doi: 10.1007/s00404-015-3725-y] [Medline: 25903520]
7. Koopmans CM, van der Tuuk K, Groen H, Doornbos JPR, de Graaf IM, van der Salm PCM, HYPITAT study group. Prediction of postpartum hemorrhage in women with gestational hypertension or mild preeclampsia at term. Acta Obstet Gynecol Scand 2014 Apr;93(4):399-407 [FREE Full text] [doi: 10.1111/aogs.12352] [Medline: 24575790]
8. Betts KS, Kisely S, Alati R. Predicting common maternal postpartum complications: leveraging health administrative data and machine learning. BJOG 2019 May;126(6):702-709. [doi: 10.1111/1471-0528.15607] [Medline: 30628159]
9. Deo RC. Machine Learning in Medicine. Circulation 2015 Nov 17;132(20):1920-1930 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.001593] [Medline: 26572668]

10. Menard MK, Main EK, Currigan SM. Executive summary of the reVITALize initiative: standardizing obstetric data definitions. Obstet Gynecol 2014 Jul;124(1):150-153. [doi: 10.1097/AOG.0000000000000322] [Medline: 24901267]
11. Venkatesh KK, Strauss RA, Grotegut CA, Heine RP, Chescheir NC, Stringer JSA, et al. Machine Learning and Statistical Models to Predict Postpartum Hemorrhage. Obstet Gynecol 2020 Apr;135(4):935-944 [FREE Full text] [doi: 10.1097/AOG.0000000000003759] [Medline: 32168227]
12. American College of Obstetricians and Gynecologists and Society for Maternal–Fetal Medicine, Menard MK, Kilpatrick S, Saade G, Hollier LM, Joseph GF, et al. Levels of maternal care. Am J Obstet Gynecol 2015 Mar;212(3):259-271. [doi: 10.1016/j.ajog.2014.12.030] [Medline: 25620372]

Abbreviations

AUROC: area under the receiver operating curve

Edited by R Kukafka; submitted 06.10.21; peer-reviewed by G Carot-Sans, J op den Buijs, L Walekr; comments to author 11.01.22; revised version received 24.02.22; accepted 13.06.22; published 18.07.22.

Please cite as:
Westcott JM, Hughes F, Liu W, Grivainis M, Hoskins I, Fenyo D
Prediction of Maternal Hemorrhage Using Machine Learning: Retrospective Cohort Study
J Med Internet Res 2022;24(7):e34108
URL: https://www.jmir.org/2022/7/e34108
doi:10.2196/34108
PMID:35849436

Original Paper

Development of an Interoperable and Easily Transferable Clinical Decision Support System Deployment Platform: System Design and Development Study

Junsang Yoo[1*], PhD; Jeonghoon Lee[2*], MD, MSc; Ji Young Min[3], MSc; Sae Won Choi[4], MD; Joon-myoung Kwon[5], MD, MSc; Insook Cho[6], PhD; Chiyeon Lim[7], PhD; Mi Young Choi[8], PhD; Won Chul Cha[1,9,10], MD, PhD

[1]Department of Digital Health, Samsung Advanced Institute for Health Science & Technology, Sungkyunkwan University, Seoul, Republic of Korea
[2]HD Junction, Inc, Seoul, Republic of Korea
[3]LINEWALKS Inc, Seoul, Republic of Korea
[4]Office of Hospital Information, Seoul National University Hospital, Seoul, Republic of Korea
[5]Medical Research Team, Medical AI Inc, Seoul, Republic of Korea
[6]Nursing Department, School of Medicine, Inha University, Incheon, Republic of Korea
[7]Department of Biostatistics, Dongguk University School of Medicine, Goyang, Republic of Korea
[8]Data Service Center, en-core Co, Ltd, Seoul, Republic of Korea
[9]Department of Emergency Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
[10]Digital Innovation Center, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
[*]these authors contributed equally

Corresponding Author:
Won Chul Cha, MD, PhD
Department of Emergency Medicine
Samsung Medical Center
Sungkyunkwan University School of Medicine
81 Irwon-ro
Gangnam-gu
Seoul, 06351
Republic of Korea
Phone: 82 3410 2053
Fax: 82 3410 0012
Email: wc.cha@samsung.com

Abstract

Background: A clinical decision support system (CDSS) is recognized as a technology that enhances clinical efficacy and safety. However, its full potential has not been realized, mainly due to clinical data standards and noninteroperable platforms.

Objective: In this paper, we introduce the common data model–based intelligent algorithm network environment (CANE) platform that supports the implementation and deployment of a CDSS.

Methods: CDSS reasoning engines, usually represented as R or Python objects, are deployed into the CANE platform and converted into C# objects. When a clinician requests CANE-based decision support in the electronic health record (EHR) system, patients' information is transformed into Health Level 7 Fast Healthcare Interoperability Resources (FHIR) format and transmitted to the CANE server inside the hospital firewall. Upon receiving the necessary data, the CANE system's modules perform the following tasks: (1) the preprocessing module converts the FHIRs into the input data required by the specific reasoning engine, (2) the reasoning engine module operates the target algorithms, (3) the integration module communicates with the other institutions' CANE systems to request and transmit a summary report to aid in decision support, and (4) creates a user interface by integrating the summary report and the results calculated by the reasoning engine.

Results: We developed a CANE system such that any algorithm implemented in the system can be directly called through the RESTful application programming interface when it is integrated with an EHR system. Eight algorithms were developed and deployed in the CANE system. Using a knowledge-based algorithm, physicians can screen patients who are prone to sepsis and obtain treatment guides for patients with sepsis with the CANE system. Further, using a nonknowledge-based algorithm, the

CANE system supports emergency physicians' clinical decisions about optimum resource allocation by predicting a patient's acuity and prognosis during triage.

Conclusions: We successfully developed a common data model–based platform that adheres to medical informatics standards and could aid artificial intelligence model deployment using R or Python.

(J Med Internet Res 2022;24(7):e37928) doi:10.2196/37928

KEYWORDS

clinical decision support system; decision making; decision aid; decision support; common data model; model; development; electronic health record; medical record; EHR; EMR; Fast Healthcare Interoperability Resource; interoperability; machine learning; clinical decision; health technology; algorithm; intelligent algorithm network; modeling

Introduction

The clinical decision support system (CDSS) is expected to play an essential role in modern medicine. The expansion of scalable data and advances in data science have led to considerable data-driven CDSS research, which has offered good opportunities to accurately reflect the clinical context with higher complexity than possible with rule-based expert systems [1]. The establishment of a research process for the development, validation, and reporting of machine-learning algorithms has made significant contributions to improving quality and reproducibility in this area [2,3].

Even the best algorithm cannot be expected to achieve its potential benefit before it is utilized in a clinical setting [4]. The transition of an algorithm from research to implementation is hindered by several factors, including social, political, economic, clinical, and technical issues [5-8]. Among these, the interoperability problem, which originates from the heterogeneity of electronic health record (EHR) systems with varying data types and structures, has been identified as an important factor that hinders CDSS implementation in a real clinical setting [9,10]. Moreover, considering that a current data-driven CDSS utilizes more variables than traditional statistical models and requires data preprocessing, it is unrealistic to expect CDSS developers to modify their model to fit each hospital's EHR system.

Dozens of standards have been introduced to overcome this interoperability issue, including the International Statistical Classification of Diseases and Related Health Problems, 10th revision; the Logical Observation Identifiers Names and Codes taxonomy; the RxNorm drug vocabulary; and the SNOMED (Systematized Nomenclature of Medicine–Clinical Terms) clinical terminology database for semantic technology integration [11]. Additionally, Health Level 7 (HL7) V2 and V3 negotiated frameworks, clinical document architectures, and HL7 Fast Healthcare Interoperability Resources (FHIR) for data exchange [12,13] have been used. The Observational Medical Outcomes Partnership (OMOP) common data model (CDM), the Sentinel CDM, and the National Patient-Centered Clinical Outcomes Research Network (PCORnet) CDM for standardized data structures and types [14,15] are other major standards developments. However, because hospitals in South Korea utilize heterogeneous home-grown EHR systems, medical informaticians face consistent difficulties in adopting international medical data standards. More recently, Clinical Quality Language (CQL) and CDS Hooks were introduced [16,17]. CQL is a language that is used in various clinical situations, including clinical decision-making, cohort definition, and clinical quality measurements. CQL can be easily integrated into HL7 FHIR via sharing functions, which helps domain experts by enhancing human readability.

The OMOP-CDM and HL7 FHIR standards are good starting points for developing a platform that can deploy an interoperable CDSS to multiple organizations [18,19]. Moreover, the OMOP-CDM has acquired the status of a de facto standard in South Korea. Over two-thirds of tertiary academic hospitals have adopted the OMOP-CDM with national research and development support [20]. Additionally, HL7 FHIR is known as a prospering standard in the medical informatics field. This standard provides a simplified data model using the FHIR 80% rule. That is, the operative guideline informally states that each resource should contain only those data elements agreed upon by 80% or more of the participants in the development effort [13,21]. Because HL7 FHIR employs a web protocol, the standard is widely used to exchange information in a variety of medical settings, including those of CDSS deployments [10,18,19].

The objective of this study was to introduce the CDM-based intelligent algorithm network environment (CANE) platform to support the implementation and deployment of a CDSS.

Methods

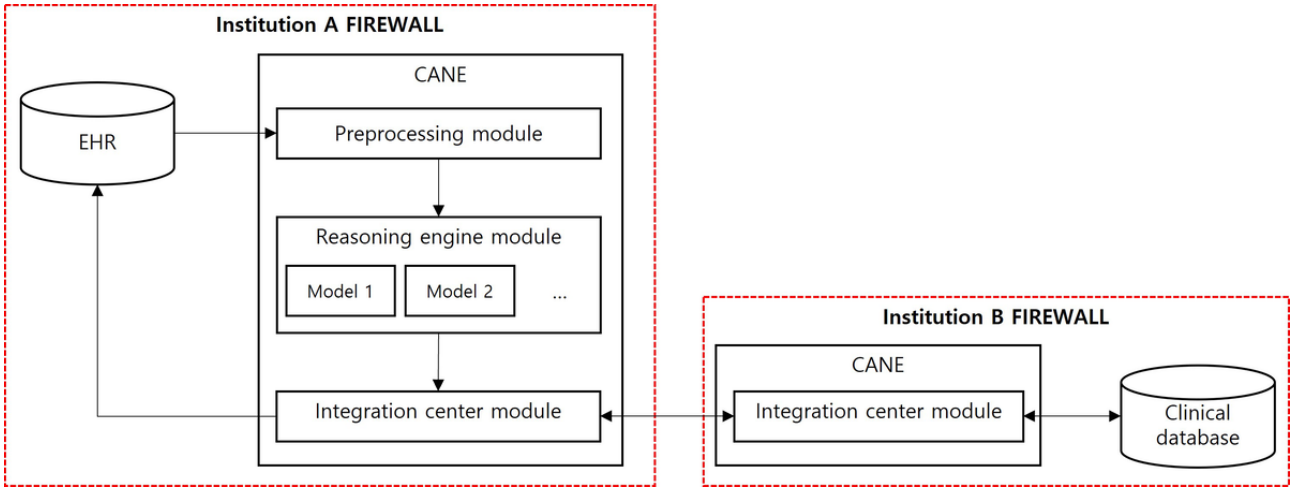
CANE Research Consortium

The CANE Research Consortium was established in May 2019 to develop a CDSS deployment platform that could extend CDSS data referencing capabilities across medical institutions. The Consortium comprised six research groups representing seven major hospitals in Seoul, Gyeonggi, and Incheon, South Korea.

CANE Architecture

The CANE platform is built on the Linux (CentOS 7.7 (1906)) operating system with 3.7-GHz octa core CPUs, 64-GB RAM, and a 2-TB hard disk drive. Microsoft.Net 5.0, MariaDB 10.4.12 (x86_64), Python 3.6.8, and Apache 2.4.6 software systems are applied. The platform consists of a preprocessing module, a reasoning engine, and an integration center module. The roles of each module were described in the CDSS operation process session (Figure 1).

Figure 1. Abstract architecture of the common data model–based intelligent algorithm network environment (CANE) platform. EHR: electronic health record.



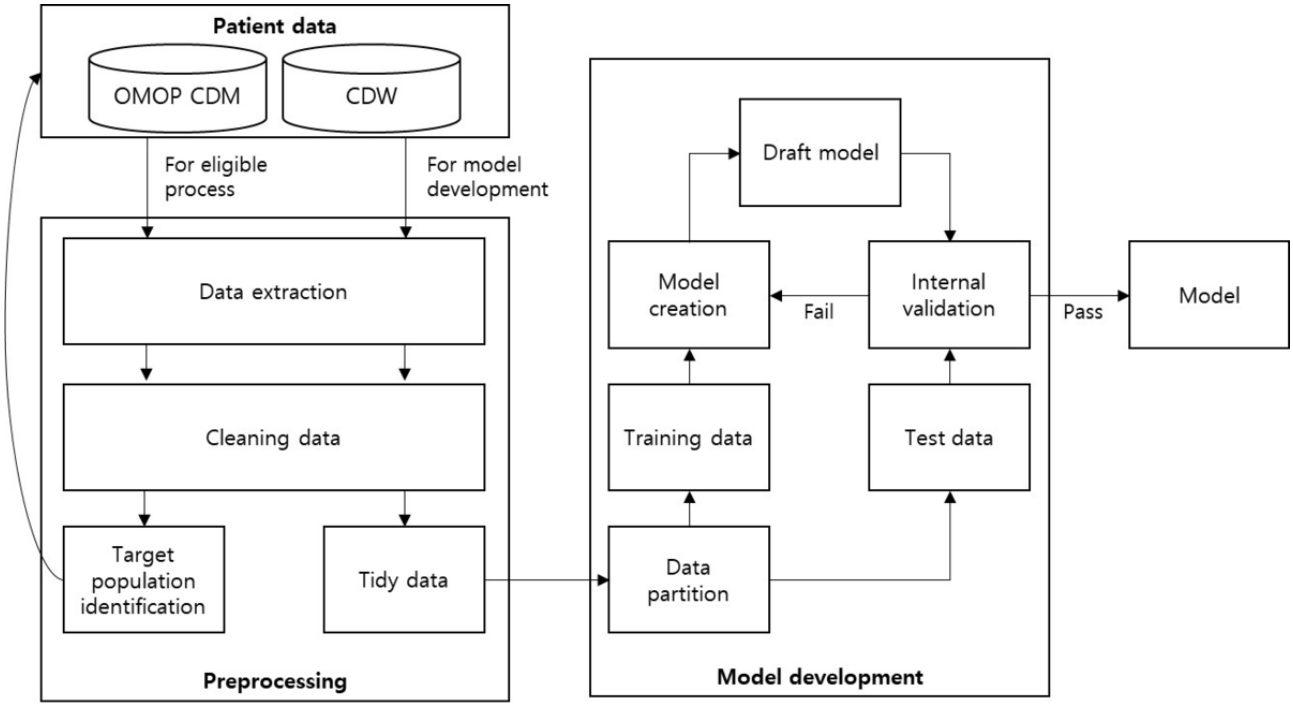
Model Development, Deployment, and Operation Processes

Phase 1: Development and Evaluation of the CDSS Reasoning Engine

Algorithms distributed in CANE are classified into knowledge- and nonknowledge-based CDSSs. A knowledge-based CDSS refers to a traditional expert system that provides informational representations of medical guidelines. A nonknowledge-based CDSS uses machine learning, a technology that recognizes patterns and makes predictions from clinical data.

Figure 2 describes the development process of the nonknowledge-based CDSS. In principle, data for algorithm development should be extracted from the OMOP-CDM database. However, learning from a local clinical data warehouse is also allowed because it may be necessary to learn from data that cannot be converted into OMOP-CDM format. There are various ways to develop a machine-learning algorithm. In this study, we followed the patient-level prediction framework: (1) target population identification, (2) predictor extraction, (3) splitting tidy data into training and test sets, (4) draft model development using a training data set, (5) iterative process of evaluating the draft model, and (6) final model confirmation.

Figure 2. Nonknowledge-based clinical decision support system model development process. CDM: common data model; CDW: clinical data warehouse; OMOP: Observational Medical Outcome Partnership.

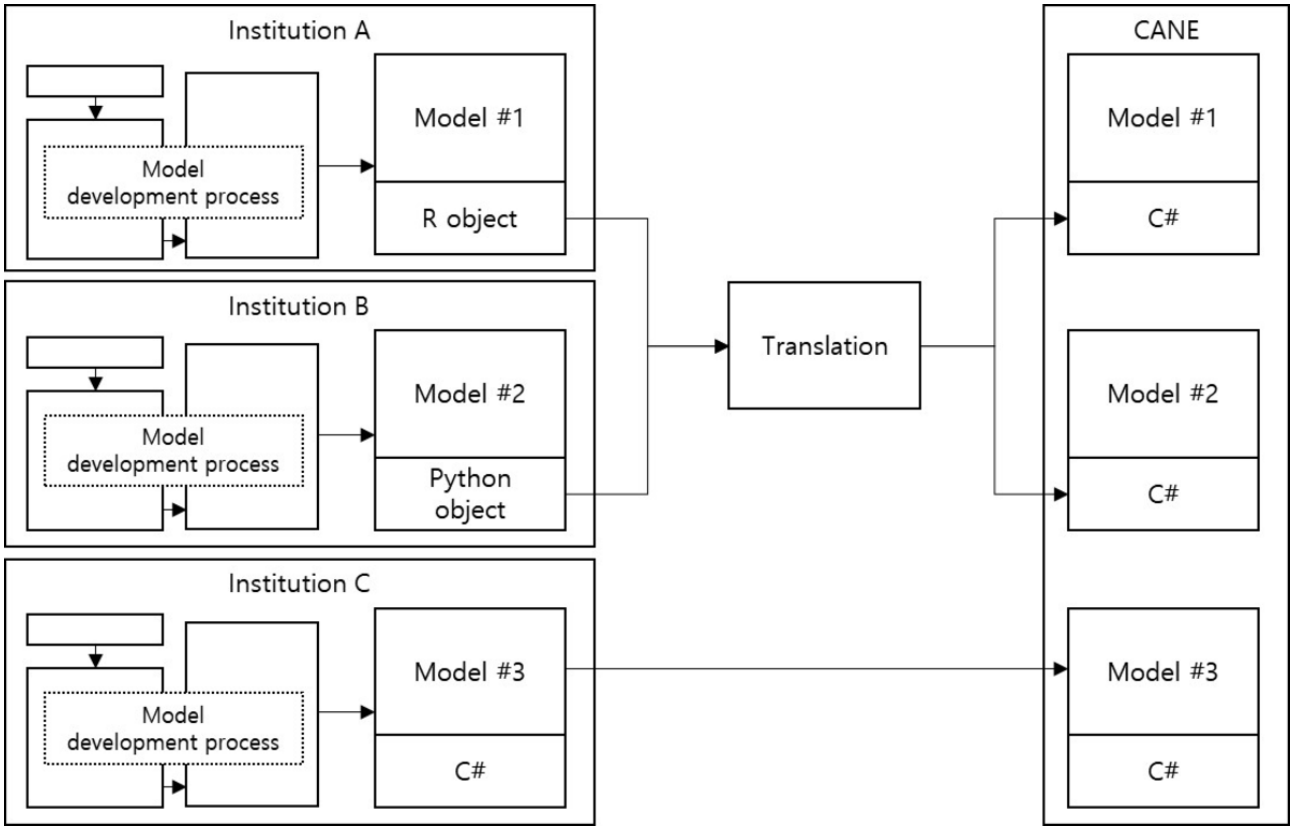


Phase 2: CDSS Deployment on the CANE Platform

The developed algorithms usually take the form of R or Python objects, which are widely used by researchers in the field of medical informatics. Converting these objects into C# language is a prerequisite for mounting the reasoning engine module of the CANE platform (Figure 3). This model conversion process is performed using the Hl.Fhir.R4 (2.0.0), Newtonsoft.Json

(12.0.3), and R.NET (1.9.0) packages. Moreover, the model formed of C# objects is simply deployed without conversion. By using C#, which is the most representative language of the .NET framework, programs can be executed on any operating system following the common language specification.

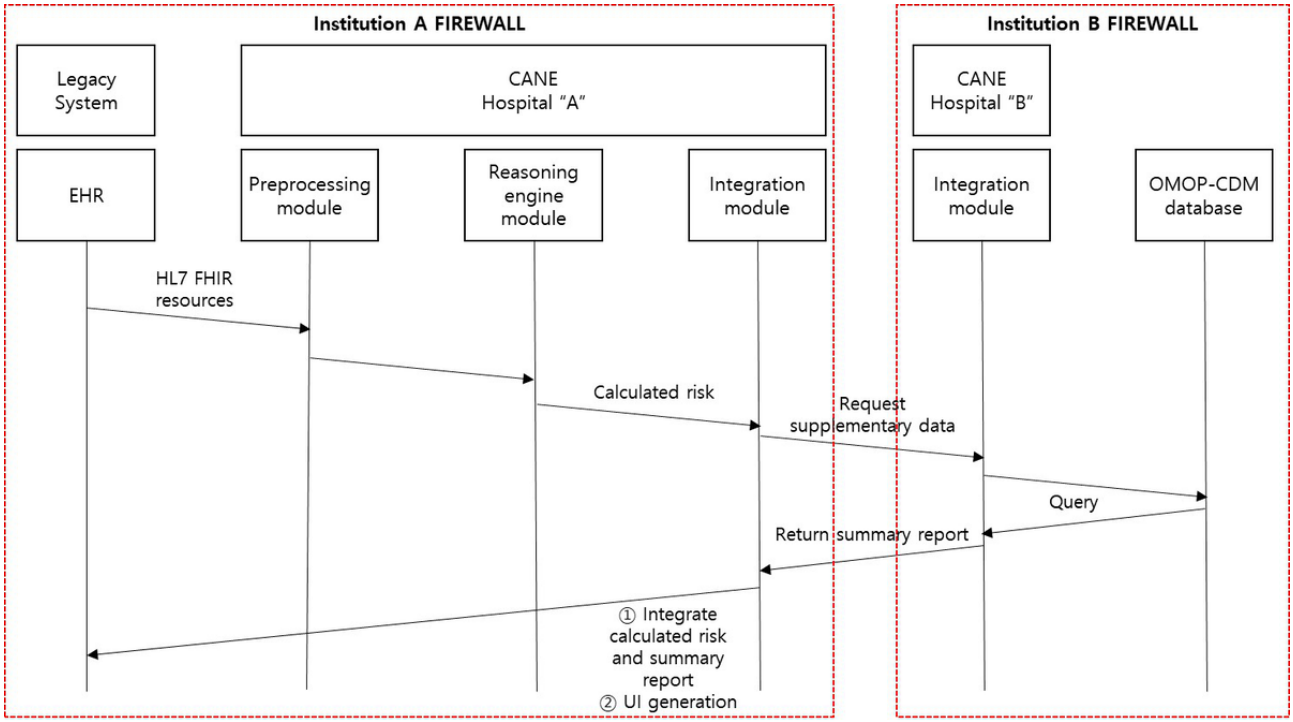
Figure 3. Model deployment process.



Phase 3: CDSS Operation Process

A user requests decision support from the CANE platform by clicking the "CANE" button on their EHR user interface (UI). Subsequently, the EHR data required for the target algorithm are converted into JavaScript Object Notation (JSON) format using HL7 FHIRs. They are then transmitted to the CANE server. Next, the CANE server parses and converts the received FHIRs to fit the target CDS algorithm. The preprocessed input data call the deployed model, and the model returns the calculated score (eg, sepsis risk score) to the integration module. All information transmissions occur between the CANE and EHR, where these systems were implemented, without integration modules. The integration module requests additional information from other institutions' CANE systems, which are interconnected via their own OMOP-CDM. To dispel potential privacy concerns, the data used for interinstitution transmission are delivered as a population-level summary rather than raw data that can be used to identify individuals. This summary data can assist the physician in possibly promoting patient behavioral changes in favorable ways. Finally, the integration module generates a UI based on the calculated score from the model and supplementary data from other institutions' CANE systems, which are then presented to the EHR. These operating processes are described in Figure 4.

Figure 4. Clinical decision support system operation process with CANE. EHR: electronic health record; FHIR: Fast Healthcare Interoperability Resource; HL7: Health Level 7; OMOP-CDM: Observational Medical Outcome Partnership-common data model; UI: user interface.



Ethics Approval

The sepsis case study was approved by the Samsung Medical Center Institutional Review Board (2019-07-034) and the emergency department (ED) case study was approved by the Sejong General Hospital Institutional Review Board (2017-1744).

Results

Overview of the CANE Platform

Currently, eight algorithms are deployed to the CANE platform; however, the research consortium intends to deploy 11 algorithms by December 2022 (Table 1). When the CANE system is integrated with an EHR system, the algorithms mounted on the CANE system can be directly called through the RESTful application programming interface (API). The CANE system also provides a web interface (Figure 5). For the web session, this study focused on the CANE platform and representative use cases of both knowledge- and nonknowledge-based CDSSs, rather than discussing all of the deployed algorithms, which would extend beyond the objectives of this paper.

Table 1. Details of the developed algorithms for clinical decision support systems (CDSSs).

Algorithm name	Objectives	Target patients
Nonknowledge-based CDSSs		
Anomaly prescription detection algorithm	To determine whether the prescription has potential information errors	Patients prescribed the following medications: heparin, Humulin, RI[a], and potassium
Test order recommendation algorithm for emergency department clinician	To recommend a prescription that is expected to need an examination based on the patient's medical record, but omitted	Patients visiting the emergency department
Triage-level decision support system	To determine the acuity of patients visiting the emergency department	Patients visiting the emergency department
Emergency department visiting patients' prognosis prediction algorithm	Algorithm for screening patients with a possibility of poor prognosis among patients visiting the emergency department	Patients visiting the emergency department
Brain injury patients' prognosis prediction algorithm	Algorithm for screening patients with a possibility of poor prognosis among patients visiting the emergency department	Patients with traumatic brain injury
Fall risk prediction algorithm	To improve patient safety by predicting patients with a high risk of falls	Inpatients
Pressure ulcer prediction algorithm	To calculate the risk of pressure ulcer, allowing for preventive action and early detection	Inpatients
Knowledge-based CDSSs		
Warfarin dosage recommendation algorithm	To recommend an appropriate dose of anticoagulant in consideration of individual patient characteristics and drug response	Patients prescribed warfarin
Insulin dosage recommendation algorithm	To recommend an appropriate dose of insulin in consideration of individual patient characteristics and response to previous insulin administration	Patients with diabetes mellitus
Dyslipidemia treatment decision support system	To integrate and represent knowledge of dyslipidemia treatment guidelines	Outpatients who require treatment for dyslipidemia
Sepsis treatment decision support system (SepsTreat)	To screen sepsis patients and provide a sepsis treatment guideline	Inpatients and patients visiting the emergency department who require treatment for sepsis

[a]RI: regular insulin.

Figure 5. CANE dashboard of the demo web page. ED: emergency department; PICC: peripherally inserted central catheter.

Use Case 1: Sepsis Treatment Decision Support System

Sepsis is a syndrome caused by infection, which is a significant public health problem that results in a patient's death without appropriate and timely treatment [22]. As the importance of early recognition and appropriate treatment is well defined, many researchers have attempted to develop early detection methods that may predict the outcomes of patients with sepsis.

We developed a sepsis treatment decision support system (SepsTreat) as one of the knowledge-based CANE algorithms (Figure 6). This algorithm is rules-based, and it provides recommended treatment guides when detecting sepsis patients.

Sepsis knowledge and treatment information are based on Sepsis-3 Guidelines [23]. Patients who show signs of organ dysfunction caused by infection are defined as sepsis patients. For SepsTreat, a sepsis patient is one whose body temperature is above 37.5°C and has been recommended a blood culture test or antibiotics. Organ dysfunction is checked using the Sepsis-related Organ Failure Assessment (SOFA) score, which cannot be calculated immediately because some score components require laboratory testing results. However, Quick SOFA (qSOFA) is a new method that helps physicians quickly assess patients suspected of infection, and offers investigative leads concerning suspected organ dysfunction. qSOFA was recently added to the Sepsis-3 Guidelines to supplement the complex SOFA score. The qSOFA score monitors only three components: systolic blood pressure, respiratory rate, and mentality.

Figure 6. Screenshot of the sepsis treatment decision support system (SepsTreat) algorithm.

A prototype of SepsTreat runs on the web as a separate instance paired with an EHR system. SepsTreat intakes component values to determine sepsis by calculating qSOFA and SOFA scores using the Glasgow Coma Scale score, systolic blood pressure, diastolic blood pressure, respiratory rate, body temperature, alveolar oxygen partial pressure, fraction of inspired oxygen, platelet count, creatinine, total bilirubin, lactic acid, blood culture order, antibiotics order, and vasopressors or inotropic medication orders. After patient examination, the physician enters each component value into SepsTreat. If the patient is in a septic condition, SepsTreat provides the recommended treatment. When the recommendation is displayed, practices that have already been completed are marked to emphasize treatments that have not yet been processed. Information retrieved from the OMOP-CDM database is also presented to help physicians manage patients. This information includes

statistics of prescribed antibiotics for sepsis patients and the statistical results of sepsis patient outcomes. Physicians in a secondary hospital or clinic can easily access tertiary information when using the CANE system by presenting statistical results from their CDM database. In the prototype version, a single center's practice statistics are presented. The final version will include patterns of prescribed antibiotics for sepsis patients from different hospitals and different outcomes.

Use Case 2: ED Patient Triage Algorithm

We incorporated a deep-learning prediction algorithm to extend CANE's flexible boundaries. Kwon et al [24] created this algorithm, which calculates triage and acuity scores for ED patients. This algorithm also predicts hospital mortality, critical care, and hospitalization metrics using information from the triage stage (eg, age, sex, chief complaint, time from symptom onset to ED visit, arrival mode, trauma, initial vital signs, and mental status). Physicians can predict patients' outcomes before the point of examination using this algorithm. From the results, physicians can deliver appropriate management before the patient's deterioration, or they may opt to hospitalize patients sooner than otherwise expected to stabilize those with severe conditions.

To integrate this algorithm into the CANE platform, we coded the algorithm's variables based on the concept ID of OMOP-CDM and retrained the algorithm using OMOP-CDM data. In the prototype of this algorithm, if a user enters input variables into the CANE platform, it presents each prediction result with a possibility and predefined risk score. Risk scores were determined in advance by researchers using statistical calculations. The resultant ED patient triage algorithm is a good example of how the CANE platform can integrate machine-learning algorithms with minimal integration effort.

Discussion

Principal Results

In this paper, we introduced the CANE platform, which supports the deployment of various types of CDSS models. Our system was primarily developed to facilitate the deployment of medical artificial intelligence (AI) algorithms developed by the CANE Consortium. However, the platform can also be used as a pipeline that integrates CDSS and hospital developers who need to collaborate on tools. Furthermore, our platform supports the transformation of algorithms developed using R and Python into C#, which was required by the CANE platform. Considering that R and Python are widely used programming languages [25], our platform could contribute to overcoming the chasm between development and deployment of medical AI.

On-Demand Intervention Type

We adopted "on-demand interventions," which is a form of noninterruptive intervention, as the approach used by the CANE system to provide decision support. According to a recent meta-analysis, interruptive intervention is the dominant form of a CDSS that has been applied and utilized in clinical practice [26]. A CDSS applied in an interruptive manner not only distorts the clinical workflow but also reports unintended consequences such as alert fatigue, which is a known factor that hinders the CDSS from achieving its purpose [4]. Hence, interruptive intervention must be applied carefully in a limited purpose [27]. Therefore, we chose on-demand intervention as the basic intervention format. By subsequently applying the CDS Hooks to the CANE system, each CDSS can readily be invoked by various intervention methods according to the clinical workflow.

Comparison With Prior Work

Despite evidence indicating that medical AI and CDSS can improve the efficacy and safety of health care delivery systems [28-30], the present situation still seems to be far from this goal. Studies have been conducted to address interoperability issues and overcome the chasm between CDSS model development and widespread deployment. Khalilia et al [18] provided convincing answers to account for this gap in terms of web services based on a service-oriented architecture. They presented a streamlined architecture that facilitates predictive modeling using OMOP-CDM structured data sets and deployed the model into a clinical workflow using HL7 FHIR. More recently, Gruendner et al [19] introduced a sophisticated and comprehensive platform that included model development, deployment, and security with a graphical UI based on OMOP-CDM and HL7 FHIR standards. Unberath et al [31] suggested an operational CDSS case to predict relapses in patients with melanoma using OMOP-CDM and the REST API.

The CANE system is distinguished from these previous works in that it provides data-driven decision support from other institutions. Thus, clinicians may refer to summary reports regarding similar situations using interconnecting integration modules at the point of care. Considering that clinicians prefer to make decisions based on peer opinions as well as CDSS information, this function is gaining importance from a behavioral science perspective [32]. Because various institutions participating in this study have constructed OMOP-CDM databases, all processes, including data queries, data analyses, and reports generation, are conducted within the hospitals' firewalls. Only summary reports are transmitted to the requesting institutions. Raw patient data are avoided via an interconnecting integration module to minimize security risks.

Model Deployment Determining Pipeline

The CANE Consortium should establish a pipeline that determines the installation of a nonknowledge-based CDSS in the CANE system, with the aim of distributing a model to other institutions. Several studies have reported that performance indicators evaluated using external data are statistically significantly lower than those evaluated using data from institutions where an AI model was developed. This could be attributed to the "Cloud of Context" issue [33]. Variations in clinical workflow, available resources, and patient characteristics among institutions hinder the generalizability of an algorithm. Hence, external validation using data from a target institution before applying AI models is an effective way to not only adjust the expectations on the model but also to prevent potential patient safety issues due to the algorithm. Therefore, we suggest that information regarding external validation must be included in the evaluation pipeline to determine whether a specific algorithm should be installed into the CANE platform.

Limitations

This study has the following limitations. The CANE platform does not embrace the machine-learning algorithms developed from the ATLAS platform, which is a widely used web-based service for building machine-learning models within the OMOP-CDM ecosystem. Second, the Consortium did not investigate the performance indices of each algorithm using either internal or external data, or the usability of the CANE platform. Further evaluation is needed in a subsequent study. Third, it is common that the performance of machine-learning algorithms differs when they are applied to other organizations. Finally, CQL and CDS Hooks, which are standards recently highlighted in the field of medical informatics, were not reflected in our system.

Conclusions

We introduced the CANE platform, which adheres to medical informatics standards (OMOP-CDM and HL7 FHIR). This system provides summary data on the treatment patterns of other institutions that could aid physicians' decision-making. Moreover, concerns regarding potential privacy issues are minimized by transmitting summary data rather than individuals' raw heath data.

Acknowledgments

This work was supported by the Technology Innovation Program (20004861, CANE for Precision CDSS) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea), and the Basic Science Research Program through the National Research Foundation of Korea, funded by the Ministry of Education (grant number 2021R1F1A1052072).

Conflicts of Interest

MYC worked for C&T, a company that developed the CANE platform. The other authors have no conflicts of interest to declare.

References

1. Farhan W, Wang Z, Huang Y, Wang S, Wang F, Jiang X. A predictive model for medical events based on contextual embedding of temporal sequences. JMIR Med Inform 2016 Nov 25;4(4):e39 [FREE Full text] [doi: 10.2196/medinform.5977] [Medline: 27888170]
2. Collins GS, Reitsma JB, Altman DG, Moons KG, members of the TRIPOD group. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. Eur Urol 2015 Jun;67(6):1142-1151 [FREE Full text] [doi: 10.1016/j.eururo.2014.11.025] [Medline: 25572824]
3. Reps J, Schuemie M, Suchard M, Ryan P, Rijnbeek P. Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. J Am Med Inform Assoc 2018 Aug 01;25(8):969-975 [FREE Full text] [doi: 10.1093/jamia/ocy032] [Medline: 29718407]
4. Ranji S, Rennke S, Wachter R. Computerised provider order entry combined with clinical decision support systems to improve medication safety: a narrative review. BMJ Qual Saf 2014 Sep;23(9):773-780 [FREE Full text] [doi: 10.1136/bmjqs-2013-002165] [Medline: 24728888]
5. Gong Y, Kang H. Usability clinical decision support. In: Berner ES, editor. Clinical decision support systems: theory and practice. Cham: Springer International Publishing; 2016:69-86.
6. Hörig H, Marincola E, Marincola F. Obstacles and opportunities in translational research. Nat Med 2005 Jul;11(7):705-708. [doi: 10.1038/nm0705-705] [Medline: 16015353]
7. Wehling M. Translational medicine: science or wishful thinking? J Transl Med 2008 Jun 17;6:31 [FREE Full text] [doi: 10.1186/1479-5876-6-31] [Medline: 18559092]
8. Glasgow R, Emmons K. How can we increase translation of research into practice? Types of evidence needed. Annu Rev Public Health 2007;28(1):413-433. [doi: 10.1146/annurev.publhealth.28.021406.144145] [Medline: 17150029]
9. Edwards A, Hollin I, Barry J, Kachnowski S. Barriers to cross--institutional health information exchange: a literature review. J Healthc Inf Manag 2010;24(3):22-34. [Medline: 20677469]
10. Hur S, Lee J, Kim T, Choi JS, Kang M, Chang DK, et al. An automated fast healthcare interoperability resources-based 12-lead electrocardiogram mobile alert system for suspected acute coronary syndrome. Yonsei Med J 2020 May;61(5):416-422 [FREE Full text] [doi: 10.3349/ymj.2020.61.5.416] [Medline: 32390365]
11. González Bernaldo de Quirós F, Otero C, Luna D. Terminology services: standard terminologies to control health vocabulary. Yearb Med Inform 2018 Aug;27(1):227-233 [FREE Full text] [doi: 10.1055/s-0038-1641200] [Medline: 29681027]
12. Mercorella M, Ciampi M, Esposito M, Esposito A, De Pietro G. An architectural model for extracting FHIR resources from CDA documents. 2016 Presented at: 2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS); November 28-December 1, 2016; Naples p. 597-603. [doi: 10.1109/sitis.2016.99]
13. Bender D, Sartipi K. HL7 FHIR: An Agile and RESTful approach to healthcare information exchange. 2013 Presented at: 26th IEEE International Symposium on Computer-Based Medical Systems; June 20-22, 2013; Porto, Portugal p. 326-331 URL: http://ieeexplore.ieee.org/document/6627810/ [doi: 10.1109/cbms.2013.6627810]

14. Hripcsak G, Duke J, Shah N, Reich C, Huser V, Schuemie M, et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. Stud Health Technol Inform 2015;216:574-578 [FREE Full text] [Medline: 26262116]
15. Collins F, Hudson K, Briggs J, Lauer M. PCORnet: turning a dream into reality. J Am Med Inform Assoc 2014;21(4):576-577 [FREE Full text] [doi: 10.1136/amiajnl-2014-002864] [Medline: 24821744]
16. Clinical Quality Language Release 1. HL7 International. 2019. URL: https://cql.hl7.org/ [accessed 2022-06-14]
17. CDS Hooks. HL7 International. URL: https://cds-hooks.org/ [accessed 2022-06-14]
18. Khalilia M, Choi M, Henderson A, Iyengar S, Braunstein M, Sun J. Clinical predictive modeling development and deployment through FHIR web services. AMIA Annu Symp Proc 2015;2015:717-726 [FREE Full text] [Medline: 26958207]
19. Gruendner J, Schwachhofer T, Sippl P, Wolf N, Erpenbeck M, Gulden C, et al. Correction: KETOS: Clinical decision support and machine learning as a service - A training and deployment platform based on Docker, OMOP-CDM, and FHIR Web Services. PLoS One 2019 Nov 13;14(11):e0225442 [FREE Full text] [doi: 10.1371/journal.pone.0225442] [Medline: 31721815]
20. Lee JH. A survey of healthcare and medical informatization in 2020 (in Korean). Sejong. 2021 Apr. URL: https://www.k-his.or.kr/board.es?mid=a10306040000&bid=0005&tag=&act=view&list_no=283 [accessed 2022-01-09]
21. Braunstein ML. FHIR. In: Braunstein ML, editor. Health informatics on FHIR: how HL7's new API is transforming healthcare. Cham: Springer International Publishing; 2018:179-203.
22. Seymour C, Kennedy J, Wang S, Chang C, Elliott C, Xu Z, et al. Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis. JAMA 2019 May 28;321(20):2003-2017 [FREE Full text] [doi: 10.1001/jama.2019.5791] [Medline: 31104070]
23. Seymour C, Liu V, Iwashyna T, Brunkhorst F, Rea T, Scherag A, et al. Assessment of clinical criteria for sepsis: for the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA 2016 Feb 23;315(8):762-774 [FREE Full text] [doi: 10.1001/jama.2016.0288] [Medline: 26903335]
24. Kwon JM, Lee Y, Lee Y, Lee S, Park H, Park J. Validation of deep-learning-based triage and acuity score using a large national dataset. PLoS One 2018;13(10):e0205836 [FREE Full text] [doi: 10.1371/journal.pone.0205836] [Medline: 30321231]
25. Meyer M. Healthcare data scientist qualifications, skills, and job focus: a content analysis of job postings. J Am Med Inform Assoc 2019 May 01;26(5):383-391 [FREE Full text] [doi: 10.1093/jamia/ocy181] [Medline: 30830169]
26. Ronan C, Crable E, Drainoni M, Walkey A. The impact of clinical decision support systems on provider behavior in the inpatient setting: a systematic review and meta-analysis. J Hosp Med 2022 May;17(5):368-383. [doi: 10.1002/jhm.12825] [Medline: 35514024]
27. Phansalkar S, van der Sijs H, Tucker A, Desai A, Bell D, Teich J, et al. Drug-drug interactions that should be non-interruptive in order to reduce alert fatigue in electronic health records. J Am Med Inform Assoc 2013 May 01;20(3):489-493 [FREE Full text] [doi: 10.1136/amiajnl-2012-001089] [Medline: 23011124]
28. 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 [FREE Full text] [doi: 10.1016/j.artmed.2019.101780] [Medline: 31980086]
29. Zheng K, Haftel H, Hirschl R, O'Reilly M, Hanauer D. Quantifying the impact of health IT implementations on clinical workflow: a new methodological perspective. J Am Med Inform Assoc 2010;17(4):454-461 [FREE Full text] [doi: 10.1136/jamia.2010.004440] [Medline: 20595314]
30. Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. J Am Med Inform Assoc 2004;11(2):104-112 [FREE Full text] [doi: 10.1197/jamia.M1471] [Medline: 14633936]
31. Unberath P, Prokosch HU, Gründner J, Erpenbeck M, Maier C, Christoph J. EHR-independent predictive decision support architecture based on OMOP. Appl Clin Inform 2020 May;11(3):399-404 [FREE Full text] [doi: 10.1055/s-0040-1710393] [Medline: 32492716]
32. Meeker D, Linder JA, Fox CR, Friedberg MW, Persell SD, Goldstein NJ, et al. Effect of behavioral interventions on inappropriate antibiotic prescribing among primary care practices: a randomized clinical trial. JAMA 2016 Feb 09;315(6):562-570 [FREE Full text] [doi: 10.1001/jama.2016.0275] [Medline: 26864410]
33. Coiera E, Ammenwerth E, Georgiou A, Magrabi F. Does health informatics have a replication crisis? J Am Med Inform Assoc 2018 Aug 01;25(8):963-968 [FREE Full text] [doi: 10.1093/jamia/ocy028] [Medline: 29669066]

Abbreviations

AI: artificial intelligence
API: application programming interface
CANE: common data model–based intelligent algorithm network environment
CDM: common data model
CDSS: clinical decision support system

CQL: Clinical Quality Language
ED: emergency department
EHR: electronic health record
FHIR: Fast Healthcare Interoperability Resource
HL7: Health-Level 7
JSON: JavaScript Object Notation
OMOP: Observational Medical Outcome Partnership
PCORnet: Patient-Centered Clinical Outcomes Research Network
qSOFA: Quick Sepsis-related Organ Failure Assessment
SepsTreat: sepsis treatment decision support system
SNOMED: Systematized Nomenclature of Medicine–Clinical Terms
SOFA: Sepsis-related Organ Failure Assessment
UI: user interface

Edited by T Leung; submitted 17.03.22; peer-reviewed by J Walsh, M Hofford; comments to author 14.04.22; revised version received 18.06.22; accepted 10.07.22; published 27.07.22.

Please cite as:
Yoo J, Lee J, Min JY, Choi SW, Kwon JM, Cho I, Lim C, Choi MY, Cha WC
Development of an Interoperable and Easily Transferable Clinical Decision Support System Deployment Platform: System Design and Development Study
J Med Internet Res 2022;24(7):e37928
URL: https://www.jmir.org/2022/7/e37928
doi:10.2196/37928
PMID:35896020

Original Paper

Effect of Virtual Reality Hypnosis on Pain Threshold and Neurophysiological and Autonomic Biomarkers in Healthy Volunteers: Prospective Randomized Crossover Study

Claire Terzulli[1,2*], MSc; Meggane Melchior[2*], PhD; Laurent Goffin[3], MSc; Sylvain Faisan[3], PhD; Coralie Gianesini[1], PhD; Denis Graff[1,4], MD; André Dufour[5], PhD; Edouard Laroche[3*], PhD; Chloé Chauvin[1,6*], MD; Pierrick Poisbeau[2*], PhD

[1]HypnoVR, Strasbourg, France

[2]Institut des Neurosciences Cellulaires et Intégratives, Centre National de la Recherche Scientifique, University of Strasbourg, Strasbourg, France

[3]ICube Laboratory, University of Strasbourg, Strasbourg, France

[4]Anesthesiology, Clinique Rhéna, Strasbourg, France

[5]Laboratoire de Neurosciences Cognitives et Adaptatives, Centre National de la Recherche Scientifique, University of Strasbourg, Strasbourg, France

[6]Department of Anesthesiology and Intensive Care, University Hospital of Strasbourg, Strasbourg, France

[*]these authors contributed equally

Corresponding Author:
Pierrick Poisbeau, PhD
Institut des Neurosciences Cellulaires et Intégratives
Centre National de la Recherche Scientifique
University of Strasbourg
8 Allée du Général Rouvillois
Strasbourg, 67000
France
Phone: 33 662710070
Email: poisbeau@inci-cnrs.unistra.fr

Abstract

Background: Virtual reality hypnosis (VRH) is a promising tool to reduce pain. However, the benefits of VRH on pain perception and on the physiological expression of pain require further investigation.

Objective: In this study, we characterized the effects of VRH on the heat pain threshold among adult healthy volunteers while monitoring several physiological and autonomic functions.

Methods: Sixty healthy volunteers were prospectively included to receive nociceptive stimulations. The first set of thermal stimuli consisted of 20 stimulations at 60°C (duration 500 milliseconds) to trigger contact heat evoked potentials (CHEPs). The second set of thermal stimuli consisted of ramps (1°C/second) to determine the heat pain threshold of the participants. Electrocardiogram, skin conductance responses, respiration rate, as well as the analgesia nociception index were also recorded throughout the experiment.

Results: Data from 58 participants were analyzed. There was a small but significant increase in pain threshold in VRH (50.19°C, SD 1.98°C) compared to that in the control condition (mean 49.45°C, SD 1.87; $P<.001$, Wilcoxon matched-pairs signed-rank test; Cohen $d=0.38$). No significant effect of VRH on CHEPs and heart rate variability parameters was observed (all $P>0.5$; n=22 and n=52, respectively). During VRH, participants exhibited a clear reduction in their autonomic sympathetic tone, as shown by the lower number of nonspecific skin conductance peak responses ($P<.001$, two-way analysis of variance; n=39) and by an increase in the analgesia nociception index ($P<.001$, paired t-test; n=40).

Conclusions: The results obtained in this study support the idea that VRH administration is effective at increasing heat pain thresholds and impacts autonomic functions among healthy volunteers. As a nonpharmacological intervention, VRH has beneficial action on acute experimental heat pain. This beneficial action will need to be evaluated for the treatment of other types of pain, including chronic pain.

(J Med Internet Res 2022;24(7):e33255) doi:10.2196/33255

KEYWORDS

virtual reality; hypnosis; pain; analgesia; autonomic changes; thermal pain; physiological; nervous system; heat pain

Introduction

Pain is an unpleasant sensory and emotional experience that is essential to the survival of living beings; however, its usefulness is lost if pain becomes chronic (duration>3 months) [1]. For this reason, chronic pain is classified as a disease by the World Health Organization (under the International Classification of Diseases 11th edition) [2]. Chronic pain is a heavy burden as it is associated with several comorbidities, ranging from emotional disturbances to severe psychosocial disorders. Furthermore, pharmacological treatments of chronic pain are often unsatisfactory, which reinforces the recommendations from pain societies across the world to suggest nonpharmacological interventions as an adjunct modality for pain management. Several of these interventions have already proven to have some efficacy in addressing various pain states, including (but not restricted to) modulation of attention, hypnosis, musicotherapy, or physical exercises [3-6]. The development of digital tools such as virtual reality (VR) portable systems represents a unique opportunity for the treatment of pain, as these tools can combine several of these nonpharmacological treatments/methods in easy-to-use eHealth solutions.

Fully immersive VR headsets isolate users from the "real world" and move them into an enjoyable alternative 3D virtual world. If well-executed, VR environments have the capability to reduce pain, as demonstrated by the pioneering work from Hoffman and collaborators [7] and recently reviewed by Chuan et al [8]. The specific mechanisms underlying this analgesic action are not fully understood, but are likely to involve several neural functions, and in particular pathways originating from the opioidergic-sensitive frontal cortex area, modulating attentional processes [6]. These projections innervate several subcortical structures known to shape the emotional responses (eg, the amygdala) and recruit the descending inhibitory control to limit sensorispinal nociceptive integration. In line with this mechanism, distraction-oriented tasks lead to activation of the amygdala that is directly correlated with a reduction of pain scores [9]. One of the rare imaging studies available [10] suggested that the activation of pain-processing structures (anterior cingulate cortex, somatosensory cortex S1, insula, thalamus) is significantly reduced by VR, which likely explains the observed reduction in pain scores after experimental noxious heat stimulations [10]. Based on these results and structures with reduced activity, it can be hypothesized that several components (ie, sensory-discriminative, affective-emotional, and cognitive) are modulated by VR. It is likely that other pathways and analgesic mechanisms remain to be discovered to explain the observed effects on pain responses.

Apart from modulation of the cortical nociceptive processing giving rise to the sensation of pain and its emotional value, pain motor responses may also be modulated in their somatic (conscious) or autonomic components. Because autonomic motor responses are less sensitive to subjective cues, they are often used in combination with other evaluation pain scales relying on subject impressions. The underlying mechanisms of VR action seem to be even more complex with the latest VR devices that often combine distracting visual cues with analgesia-promoting auditory sensory stimulations, ranging from passive listening of music to hypnotic suggestions [3,4,11]. Altogether, VR is likely to be an interesting tool to reduce pain and its unpleasantness.

In this study, we used virtual reality hypnosis (VRH), which combines the computer-generated immersive environment of VR with a hypnotic script [12]. Similar to classical hypnosis treatments, the VRH scheme used in this study consists of an "induction period" (usually an invitation to focus one's attention) and a "dissociation" period (separation between the mental and the environment), followed by suggestions of pain reduction as analgesia. Although the beneficial effect of hypnosis on pain has been previously demonstrated [8,13], only a few studies have investigated VRH effects on pain levels [14-17]. Here, we used nociceptive heat stimulations triggered by an ultrafast thermode (TCS-II, QST-Lab; 300°C/second) to evaluate the effects of VRH (HypnoVR) on pain threshold and on contact heat evoked potentials (CHEPs).

Our hypothesis was that this VRH device, combining VR and hypnotic suggestions, increases pain thresholds, an effect that can possibly be predicted by cortical electrophysiological signatures and autonomic monitoring. Thus, the secondary objectives consisted of analyzing VRH-associated changes of several physiological biomarkers such as heart rate and heart rate variability, analgesia nociception index, respiratory rate, and skin conductance responses (SCRs) with and without VRH.

Methods

Participants

Sixty adult participants were included in the study (32 women and 28 men). Participants had to be affiliated with the French social security system, and could not participate if they had unbalanced epilepsy, psychotic disorders, depression, hearing and/or visual impairments preventing the use of VRH, or chronic diseases that may influence pain perception (eg, chronic pain, diabetes); if they were participating in another clinical study; were unable to provide informed consent; or refused to participate. Women could not participate if they were pregnant or breastfeeding.

Ethics Considerations

This study was approved by the ethics review board of CPP OUEST IV-Nantes (approval date March 28, 2019; Agence Nationale de Sécurité du Médicament et des Produits de Santé, French Ministry of Health, information date May 7, 2019; IdRCB n° 2018-A02992-53). All participants signed a written informed consent form prior to participation.

Study Design

This was an open, single-center, comparative, crossover study. Each participant performed the experiment in both the VRH and control (without VRH) conditions. To limit a potential order

effect, the time of VRH application was randomly counterbalanced across participants. At the time of inclusion, suggestibility was assessed using the standardized Barber scale test [18] and the anxiety trait was assessed using the State-Trait Anxiety Inventory (STAI) self-administered questionnaire [19]. During the experiment, participants were exposed to the different thermal stimulation protocols during the control or VRH condition. Cardiac frequency, respiration rate, electrodermal conduction, and analgesia nociception index (ANI; MDoloris, France) were recorded throughout all experiments. Occurrence of adverse events was also recorded.

VRH Setup

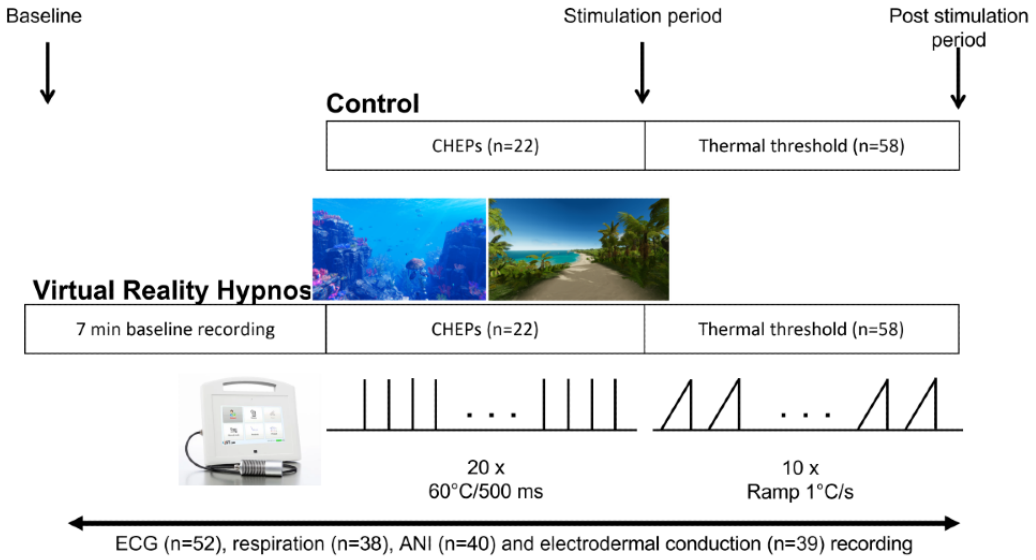
For the experiment, the VR headset was an Oculus Rift (resolution: 1080×1200 pixels per eye; field of view: 110°; frame rate: 90 Hz) coupled to a laptop computer (Asus GL502VS managed by an IntelCore i7-6700HQ processor at 2.6 GHz; RAM: 16 GB; graphics card: Nvidia GeForce GTX 1070; Windows 10 64-bit). The sound was delivered by the Oculus Rift headset.

VRH was delivered through the HypnoVR application, coupling 3D immersive and dynamic visual scenery (walking on a beach or scuba diving) with a standardized prerecorded hypnotic script (including relaxing and analgesic suggestions available in several languages) as well as musical background following music therapy principles. The hypnotic script was the same for all environmental scenarios. The musical piece followed a U-shape sequence so that, together with visual experiences, it progressively helped the participants reach a state of cardiac coherence [20]. Participants were given the opportunity to choose between one of the two preferred visual sceneries, a male or female voice for hypnotic suggestions, and musical background among three melodies. The VRH session lasted 20 minutes. The first 3 minutes correspond to the "induction period," a sequence aiming to focus an individual's attention with deep-breathing exercises. This was followed by well-being and pain relief suggestions (including changing pain sensations into something else, reduction in pain, optimization of well-being, changes in focus attention away from pain, and increased ability to ignore pain). The session ended with a 2-minute sequence leading the participant back to a normal conscious state.

Thermal Stimuli and Experimental Protocol

Thermal hot stimuli were applied with a thermal stimulator (TCS II, QST.Lab, Strasbourg, France) following the general scheme indicated in Figure 1. The thermode was placed on the dorsal surface of the nondominant hand. The first stimulation session was used to evaluate CHEPs, comprising 20 transient stimulations at 60°C/500 milliseconds. The temperature increase rate was set at 300°C/second and the decrease rate was set at 200°C/second. The baseline temperature was set at 30°C. To avoid increased sensibility of the stimulated area, the thermode was moved from one contact field stimulation to another (always on the dorsal part of the nondominant hand).

Figure 1. Overview of the experimental protocol indicating 3 periods: virtual reality hypnosis induction, stimulation protocols with acute heat stimulation (for somatosensory event-related potential measures), and temperature ramps (for pain threshold determination). Arrows indicate the period of measurements of autonomic parameters. Two representative images of the virtual environment proposed to patients are shown. n corresponds to the number of subjects included in each analysis. ANI: analgesia nociception index; CHEP: contact heat evoked potential; ECG: electrocardiogram.



This session was followed by a second stimulation session used to evaluate heat pain thresholds, which was performed using the limit method with 10 ascending ramps. The temperature was increased at a rate of 1°C/second, from skin temperature (measured before the first ramp) to a temperature that volunteers considered as painful. Volunteers were given the instruction to stop the temperature increments with a push button when they felt that the stimulation was becoming painful. The pain threshold was assessed by averaging 10 trials. To take into account interindividual differences in skin temperature, the absolute pain threshold was measured as well as the difference between the skin temperature (measured just before the first ramp) and the absolute value of the threshold (Δ temperature). All participants were subjected to the hot stimulation sequence, enabling measurements of both CHEPs and pain thresholds with and without VRH. Stimuli occurred after a few minutes (maximum 7 minutes) of rest.

Physiological Data

Cortical CHEPs (from electroencephalography [EEG] data) were recorded using Active Two AD-Box coupled with a 32–active electrodes cap respecting the 10/20 system (Biosemi). Ground electrodes (common mode sense and a driven right leg) were located between C3-Cz and between Cz-C4, respectively. The sampling frequency was set to 2048 Hz. EEG data were collected and monitored throughout the recording with Actiview version 8.0 (BioSemi B.V., WG-Plein 129, 1054SC). Raw data were preprocessed (offline) with Cartool software [21] with the help of a 50-Hz notch filter combined to 0.1-Hz high- and 80-Hz low-pass filters. The EEG file was then segmented into discrete single-trial epochs, 1500 milliseconds long with a 500-millisecond baseline before stimulus onset and a 1000-millisecond poststimulus period. Successful trials were averaged for each participant. Movement and eye-blink artifacts were removed manually. Data were checked individually for flat or noisy periods with Cartool software before further analysis.

Physiological parameters were recorded with BIOPAC MP150 (BIOPAC System Inc). The electrocardiogram (ECG; beats/minute) was measured with BIOPAC ECG100C. The breathing rate (cycles/minute) was measured by a thermistor that determined the difference in temperature between inhaled and exhaled air (BIOPAC TSD202F). SCRs were measured from the extremities of the index and middle fingers of the dominant hand (BIOPAC TSD203). Acquisition was performed through a homemade software collecting and synchronizing data from the BIOSEMI and BIOPAC acquisition equipment. The sampling rate was set to 500 Hz for BIOPAC signals. Preprocessing of ECG included a 5th-order Butterworth high-pass filter of 0.5 Hz and a 50-Hz Notch filter. The detection of R-R peak intervals enabled extracting the percentage of successive R-R intervals that differ by more than 50 milliseconds (pNN50) and the root mean square of successive R-R interval differences (RMSSD) was calculated. A 3rd-order finite impulse response filter was applied to remove the electrical noise for SCR. A simple pic detection was performed for respiration data with no additional data treatment. ANI scores were calculated by the mDoloris monitor (MetroDoloris) following a previously reported method [22].

ECG, respiration, and SCR (all synchronized within one file) were epoched in a 1-minute-long file. Epoched data were analyzed with Clampfit (Molecular Devices) and Python 3.8 (especially the Neurokit library [23]).

Data Collection

Data were prospectively collected using an audit form established for the study. All personal identifying information was removed from the database in accordance with regulations prescribed by the French data protection authority Commission Nationale de l'Informatique et des Libertés (CNIL 2213128). Collected data included the demographic characteristics (age, sex, education level) and if the participants had previously experienced motion sickness, as it might be a risk factor for nausea during VRH.

Statistical Analysis

Results are expressed as mean (SD). The statistical analyses included a descriptive component and an analytical component. All statistical analyses were performed with GraphPad Prism software (version 6). The significance level was set at α=.05 for all analyses. Normality of the distributions was tested using the Shapiro-Wilk normality test. Differences between male and female participants in baseline characteristics were analyzed using the Mann-Whitney *U* test, unpaired *t*-test, or Fisher exact test, as applicable. Cohen *d*, which specifically measures the effect size of the difference between two means, was calculated as (mean 1–mean 2)/pooled SD for both groups; thresholds of 0.3, 0.5, and 0.8 were considered as a small, medium, and large effect size, respectively. Differences in thermal sensitivity between the control and VRH conditions were assessed using the Wilcoxon matched-pairs signed-rank test. CHEPs were analyzed with a Student *t*-test for paired data and with two-way analysis of variance (ANOVA) followed by the Tukey posthoc test. Heart rate, respiration rate, and skin conductance data were analyzed with two-way ANOVA followed by posthoc Dunn-Sidak multiple-comparison correction. Statistical significance of the heart rate variability parameters, pNN50, and RMSSD was assessed using a Wilcoxon matched-pairs signed-rank test. ANI was analyzed with a paired *t*-test. To assess the effect of VRH, results were compared with the control condition. Note that the number of analyzed participants varies for different analyses because of unexpected and random electrical artifacts appearing in the periods of interest (ie, around the trigger time). If such artifacts occurred, we only retained uncontaminated signals to perform the analysis for a given participant.

Results

Participant Characteristics

A total of 60 participants fulfilled the inclusion criteria and were included in the study. Two participants (1 man and 1 woman) were excluded owing to incomplete data. Thus, 58 participants were included in the final analysis (Figure 2), including 27 men and 31 women. The mean age was 30 years, which ranged from 19 to 56 years (Table 1). There were no significant age or sex correlations with pain threshold measures, CHEPs, or autonomic parameters.

Most participants (40/58, 69%) did not previously suffer from travel sickness. The Barber suggestibility scores ranged from 0 to 7 out of a possible total of 8, with a mean of 3.2. Men and women had similar Barber suggestibility scores, demographic characteristics, and baseline measured variables, except for the STAI score and history of motion sickness, which were slightly higher in women. No adverse events were reported by any subject during the study.

Figure 2. Flow diagram of screened, randomized, and excluded participants.

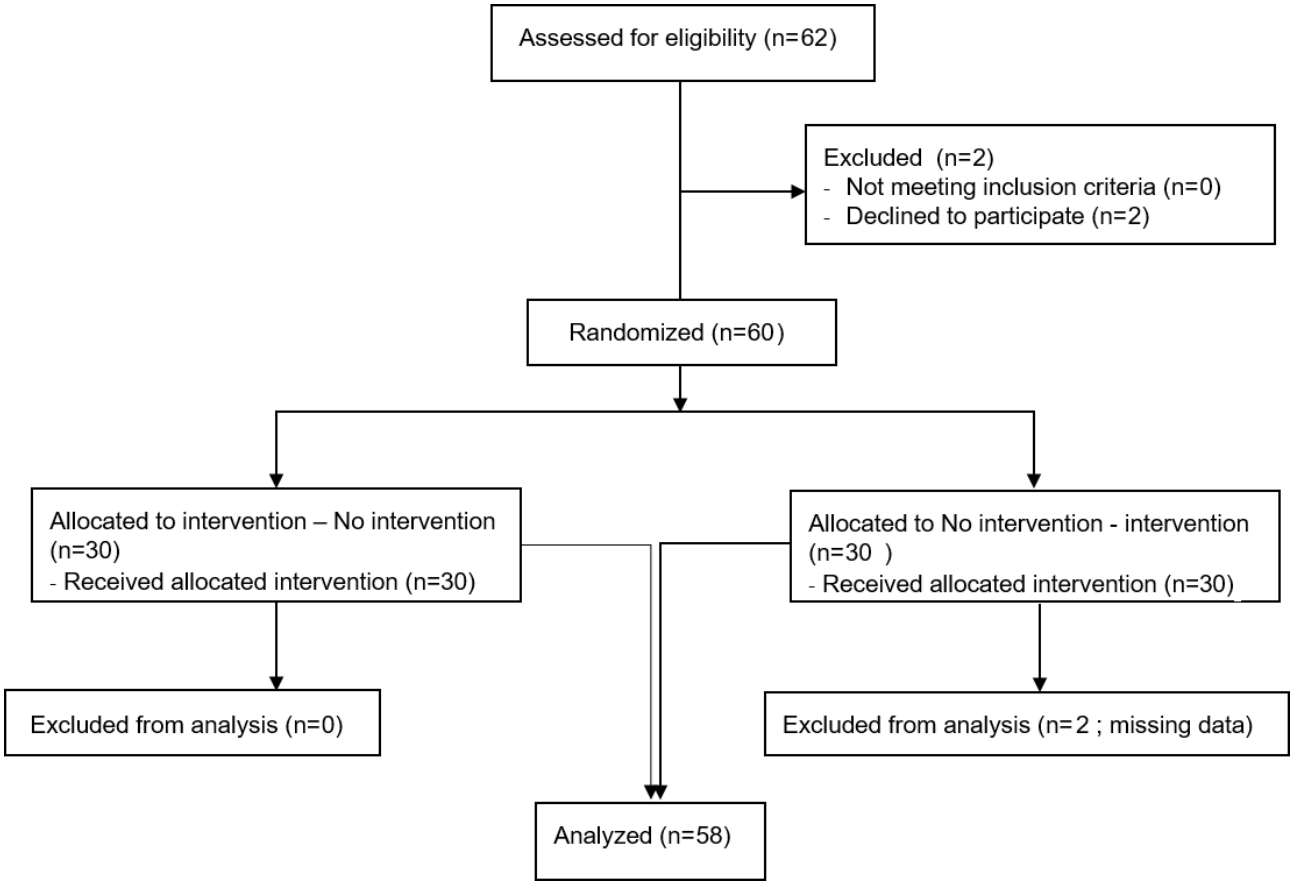


Table 1. Participant characteristics.

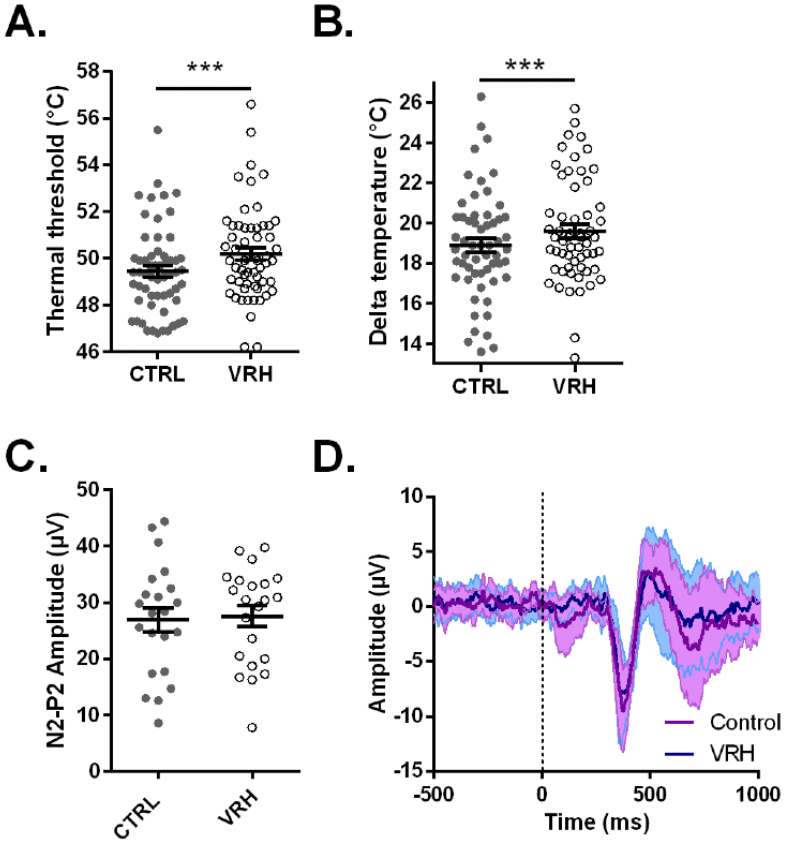
Characteristics	All participants (N=58)	Range (Total)	Women (n=31)	Men (n=27)	*P* value (women vs men)
Age (years), mean (SD)	30 (9.4)	19-56	31 (10.4)	29 (8.1)	.66[a]
Education (years postbac[b]), mean (SD)	4.3 (2.3)	1-8	3.8 (2.6)	4.8 (1.9)	.08[a]
STAI[c] (score/80), mean (SD)	37.8 (9.2)	23-60	39.9 (8.6)	35.3 (9.4)	.03[a]
Barber (score/8), mean (SD)	3.2 (1.6)	0-7	3.3 (1.6)	3.1 (1.6)	.65[d]
Travel sickness history, n	18	N/A[e]	15	3	.004[f]
BMI, mean (SD)	22.5 (7.0)	18 – 32.4	21.9 (2.7)	23.2 (3.3)	.13[a]

[a]Mann-Whitney *U* test.
[b]After undergraduate college degree.
[c]STAI: State-Trait Anxiety Inventory.
[d]Unpaired *t*-test.
[e]N/A: not applicable.
[f]Fisher exact test.

VRH Effect on Pain Thresholds and Pain Evoked Potentials

As illustrated in Figure 3A and B, we observed significantly higher mean temperature thresholds (49.45°C, SD 1.87 in the control and 50.19°C, SD 1.98 in the VRH group; Wilcoxon matched-pairs signed-rank test $P<.001$; N=58) and Δ temperature (18.91°C, SD 2.65 in the control and 19.59°C, SD 2.58 in the VRH group; Wilcoxon matched-pairs signed-rank test $P<.001$; N=58) in the VRH compared to control condition. The mean absolute difference between the control and VRH conditions was 0.74°C (95% CI 0.43-1.06), and the mean difference in Δ temperatures was 0.68°C (95% CI 0.28-1.08). Cohen *d* calculation yielded a small effect size of 0.384.

Figure 3. Effect of virtual reality hypnosis (VRH) on pain thresholds and somatosensory event-related potentials. A. Mean (SEM) absolute temperature before and after VRH. B. Mean (SEM) delta temperature (ie, the difference between the temperature threshold and skin temperature) in both conditions. C. Evolution of the mean (SEM) amplitude of N2-P2 (in µV) between control and VRH conditions. D. Superimposed mean traces of somatosensory event-related potentials obtained during VRH and without VRH, represented with their respective SDs. ***P<.001; Wilcoxon matched-pairs rank test (N=58). CTRL: control group, without VRH.



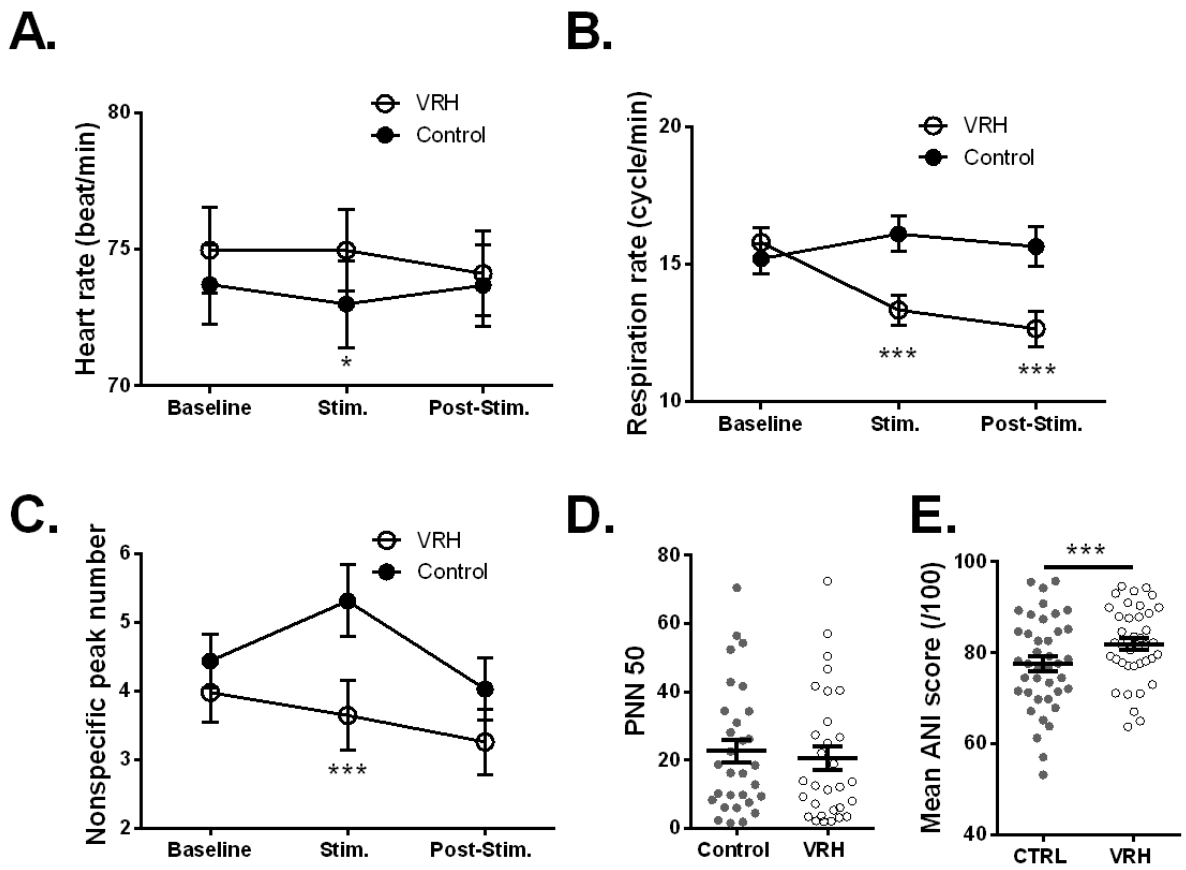
We then measured CHEPs during the control and VRH conditions. Mean epoch traces for all participants with more than two successful trials in both conditions were retained in the analysis and are shown in Figure 3D (mean 3.5, SD 1.8 successful trials; range 2-8). The mean N2-P2 amplitude was 26.9 (SD 9.92) µV and 27.6 (SD 8.66) µV in the control and VRH condition, respectively (Figure 3C). No significant difference in the N2-P2 amplitude between the two conditions was found (mean of differences 0.67, 95% CI –1.59 to 2.94; t_{21}=0.62, P=.54; n=22).

VRH Modulation of Pain-Induced Physiological Changes

Mean values for heart rate, respiration rate, nonspecific SCR, heart rate variability parameters (ie, pNN50, RMSSD, and ANI) are shown in Figure 4 in various conditions.

The mean heart rate remained globally stable during the entire protocol, which was divided into a baseline period, a period corresponding to the intervals between the stimulation protocols, and a period occurring after the stimulation protocols. No differences were observed between the control and VRH conditions for each period (Figure 4A; two-way ANOVA time×condition $F_{2, 102}$=1.31, P=.27; n=52). The respiration rate was similar at baseline for both conditions. No change in the respiratory rate was observed during the three phases of the analysis for the control group, whereas a significant decrease in this rate was observed in the VRH condition during the stimulation period and shortly after, at the end of the recording period (Figure 4B; two-way ANOVA time×condition $F_{2,74}$=13.71, P<.001; n=38). As for the other parameters, SCR was similar at baseline for both conditions, although the number of nonspecific peaks increased during the stimulation period in the control condition. This increase was not observed in the VRH condition and mean values remained stable (Figure 4C; two-way ANOVA condition $F_{1,38}$=13.74, P<.001; n=39). The mean value for the heart rate variability parameter pNN50 remained similar in the VRH and control conditions (Figure 4D; Wilcoxon matched-pairs signed-rank test P=.12; n=40). No change in RMSSD was observed (Wilcoxon test P=.96, n=40; control mean 36.81, SD 24.32; VRH mean 37.25, SD 20.56). This was not the case for ANI mean values, which significantly increased in the VRH condition compared to the control (Figure 4F; paired t-test t_{39}=3.76, P<.001; n=40). Interestingly, the ANI score is calculated by using the ratio between low-frequency parasympathetic and high-frequency sympathetic frequency powers. This parameter has been shown to be more sensitive in pain states and discomfort [20].

Figure 4. Effect of virtual reality hypnosis (VRH) on autonomic parameters (mean, SEM). Mean heart rate (A), respiration rate (B), and nonspecific skin peak conductance (C) at baseline, during stimulation (Stim.; ie, between somatosensory event-related potential stimulations and ramps), and after the last stimulation (Post-Stim.) for the control and VRH conditions. D. Percentage of successive R-R intervals that differ by more than 50 milliseconds (PNN50) as an index of cardiac variability. E. Analgesia nociception index (ANI) for both conditions. *P<.10, ***P<.001 with Sidak multiple comparison test for panels A to C; ***P<.001 with paired t-test for panel E.



Discussion

This study investigated the effect of VRH on pain thresholds and CHEPs in response to heat stimuli in healthy adult volunteers (men and women). Changes in several physiological parameters were also monitored during the stimulation protocol. We found that VRH increased the heat pain threshold, reduced the mean respiratory rate during the VRH session, and increased the ratio between parasympathetic and sympathetic tones, as seen by the stability of SCR and the increase in ANI score.

Our findings are consistent with a recently published study, in which the authors measured the effect of VR on heat-pain tolerance limits [24]. They tested two VR conditions: (1) an immersive condition in which the viewer could experience a 360° video and audio immersion, and (2) a nonimmersive condition with audio and 2D video only. They found an increase in the pain tolerance threshold in both conditions, which was higher with the immersive VR (by approximately 1°C). The effect size on pain threshold for our study and this previous work of Colloca et al [24] was small and of similar amplitude (Cohen *d* 0.384 and 0.321, respectively). In addition, another study showed a decrease in worst pain intensity and pain unpleasantness following VR in response to thermal stimulation of the foot [25]. Interestingly, the increase in pain threshold reported by healthy volunteers in this study was not associated with changes in CHEPs amplitude. CHEPs result from the activation of different cortical structures, including the somatosensory, insular, and cingulate cortices [26], but are also modulated by attentional processes and stimulus salience [27]. This likely suggests that the cortical processing of the heat stimulus was not modulated by VRH. One possible explanation is that VRH did not sufficiently distract the participant's attention to affect the CHEPs amplitude. However, VRH might still impact the activity of subcortical structures involved in the modulation of pain and in its perception as a negative emotion [6,28,29]. This working hypothesis is in line with a recent study showing that active VR (eg, a game) but not passive VR (eg, a movie) decreases brain activity following painful electrical stimuli, which was associated with a reduction in the experience of pain [30]. In this study, we did not observe any effect of VRH on pain scores despite a significant difference in heat pain threshold. Associating the measurement of pain sensation with brain imaging could provide further information on the mechanisms underlying the effects of VRH on pain. VR (but not VRH) effects were investigated in one study using functional magnetic resonance imaging brain scans, in which reductions in the activity of key structures of the pain matrix were observed [10].

Taking advantage of the simultaneous recording of some physiological parameters under VRH, we could observe significant changes of certain signals even though the nociceptive stimulations were of short duration. We observed a decrease in respiration rate and in SCR in VRH compared to the control condition, which confirmed the efficiency of the VRH script to promote relaxation and a possible decreased anxiety level that is known to reduce pain [31]. SCR peaks, likely reflecting sympathetic activation and arousal, were also lowered in our study [32]. This effect may also account for a reduction in arousal contributing to the effect observed on pain thresholds [33]. This finding is in line with the elevation of the ANI score, which is used to monitor the comfort (ie, parasympathetic tone). However, we failed to detect any effect in temporal-domain heart rate variability parameters. This lack of effect was also reported in a recent study, where the SD from normal to normal was affected only in one VR condition (immersive Ocean) but not in the others [24].

Compared to other VR devices, the VRH device used in this study includes not only visual and auditory immersive clues but also a hypnotic script following the classical hypnosis sequence for treatment purpose (ie, induction, dissociation, and suggestions of pain reduction). Hypnosis is an active cognitive treatment that allows the mind to influence sensations and perceptions of the body [30]. Accordingly, the interaction between the patient and the therapist aims to engage the patient in cognitive processes to reduce pain. The efficacy of VRH has also been demonstrated in patients suffering from traumatic pain or burn pain [15,34] and a case report suggested a positive effect on neuropathic pain [14].

The main limitation of this study is linked to the experimental setup, as the hardware/software connections generated electrical artifacts in some cases so that the data from several participants could not be properly analyzed and were thus withdrawn from a specific analysis. Another limitation concerns the characteristics of the enrolled participants, who were healthy volunteers, highly educated, and of young age, which is not representative of the general population [35-37]. Finally, we measured responses to heat stimulations applied with a quantitative sensory testing apparatus in a nonstressful laboratory environment, which does not correspond to the clinical reality of pain, regardless of whether it is acute or chronic. For example, procedural pain also implies that the individual is experiencing other forms of stress and discomfort linked to olfactory, auditory, or visual cues induced by the medical procedure. In this context, VRH might be even more beneficial by removing patients from these cues, redirecting their attention to a more pleasant environment with analgesic hypnotic suggestions [3,11]. With regard to chronic pain states, which are often associated with emotional comorbidities, it would be of interest to study the effect of repeated VRH sessions when the device is freely accessible to patients (and not only the effect of a single VRH session as performed in this study). This will enable recording the improvement in subjective pain score (intensity and unpleasantness) over time as well as the impact on quality of life.

Collectively, the results of this study suggest that VRH has a small but significant beneficial effect on acute heat pain. This effect of VRH may involve multiple modulatory pathways, modifying the perception of pain and its expression through conscious and autonomic parameters, all leading to a better relaxation state. Acute or repeated use of VR might hence provide therapeutic benefits in patients suffering from pain, including when they are outside a hospital structure, as found in recent studies [38,39]. This will require further investigations; however, compared to other nonpharmacological interventions, VRH has the advantage of being easy to use and available at home for repeated use (after a short session of therapeutic education) without the need of any medical assistance.

Acknowledgments

We thank the following research programs of excellence for their support: French National Research Agency (ANR) through the Programme d'Investissement d'Avenir (contract ANR-17-EURE-0022, EURIDOL Graduate School of Pain). PP received financial support from Institut Universitaire de France as a senior fellow. We thank Michel David for help with implementation of the research, along with Paolo Loureiro de Sousa and the staff of the "Imagerie, Robotique et Innovation pour la Santé (IRIS) "ICube" platform, Bâtiment "IRM 3T" of Strasbourg for their cooperation. We also thank Professor Kristian Torp for his thorough review. The project was supported by "Grand Est" regional funding and FEDER as part of the 2017 call for expressions of interest "Digital Economy." The following program and institutional partners also contributed from the Centre National de la Recherche Scientifique, Université de Strasbourg. CT is supported by a PhD scholarship from the Conventions Industrielles de Formation par la Recherche program coordinated by the Association Nationale pour la Recherche et la Technologie.

Authors' Contributions

Study conception: CC, DG, EL, SF, LG; Data acquisitions and analysis: CT, LG, SF, AD, PP, MM; Article writing: CT, MM, CC, PP; Funding acquisition: EL, CC, PP, SF, DG.

Conflicts of Interest

CC and DG hold a pending patent for the software HypnoVR used in this study. AD holds a pending patent for the thermal stimulator used in this study. The other authors report no conflict of interest related to this work.

References

1. Raja SN, Carr DB, Cohen M, Finnerup NB, Flor H, Gibson S, et al. The revised International Association for the Study of Pain definition of pain: concepts, challenges, and compromises. Pain 2020 Sep 01;161(9):1976-1982 [FREE Full text] [doi: 10.1097/j.pain.0000000000001939] [Medline: 32694387]
2. Bennett MI, Kaasa S, Barke A, Korwisi B, Rief W, Treede R, IASP Taskforce for the Classification of Chronic Pain. The IASP classification of chronic pain for ICD-11: chronic cancer-related pain. Pain 2019 Jan;160(1):38-44. [doi: 10.1097/j.pain.0000000000001363] [Medline: 30586069]
3. Jensen MP. Hypnosis for chronic pain management: a new hope. Pain 2009 Dec;146(3):235-237. [doi: 10.1016/j.pain.2009.06.027] [Medline: 19596518]
4. Lunde SJ, Vuust P, Garza-Villarreal EA, Vase L. Music-induced analgesia: how does music relieve pain? Pain 2019 May;160(5):989-993. [doi: 10.1097/j.pain.0000000000001452] [Medline: 30507782]
5. Kroll HR. Exercise therapy for chronic pain. Phys Med Rehabil Clin N Am 2015 May;26(2):263-281. [doi: 10.1016/j.pmr.2014.12.007] [Medline: 25952064]
6. Villemure C, Bushnell CM. Cognitive modulation of pain: how do attention and emotion influence pain processing? Pain 2002 Feb;95(3):195-199. [doi: 10.1016/S0304-3959(02)00007-6] [Medline: 11839418]
7. Hoffman HG, Doctor JN, Patterson DR, Carrougher GJ, Furness TA. Virtual reality as an adjunctive pain control during burn wound care in adolescent patients. Pain 2000 Mar;85(1-2):305-309. [doi: 10.1016/s0304-3959(99)00275-4] [Medline: 10692634]
8. Sakai H. Evaluation of corrosion rates of dental Ag-based alloys by coulostatic method. Shika Zairyo Kikai 1988 Sep;7(5):769-778. [Medline: 3272030]
9. Tracey I, Ploghaus A, Gati JS, Clare S, Smith S, Menon RS, et al. Imaging attentional modulation of pain in the periaqueductal gray in humans. J Neurosci 2002 Apr 01;22(7):2748-2752 [FREE Full text] [Medline: 11923440]
10. Hoffman HG, Richards TL, Coda B, Bills AR, Blough D, Richards AL, et al. Modulation of thermal pain-related brain activity with virtual reality: evidence from fMRI. Neuroreport 2004 Jun 07;15(8):1245-1248. [doi: 10.1097/01.wnr.0000127826.73576.91] [Medline: 15167542]
11. Jensen MP, Patterson DR. Hypnotic approaches for chronic pain management: clinical implications of recent research findings. Am Psychol 2014;69(2):167-177 [FREE Full text] [doi: 10.1037/a0035644] [Medline: 24547802]
12. Askay SW, Patterson DR, Sharar SR. Virtual reality methods. Contemp Hypn 2009 Mar;26(1):40-47 [FREE Full text] [doi: 10.1002/ch.371] [Medline: 20737029]
13. Thompson T, Terhune DB, Oram C, Sharangparni J, Rouf R, Solmi M, et al. The effectiveness of hypnosis for pain relief: A systematic review and meta-analysis of 85 controlled experimental trials. Neurosci Biobehav Rev 2019 Apr;99:298-310. [doi: 10.1016/j.neubiorev.2019.02.013] [Medline: 30790634]
14. Oneal BJ, Patterson DR, Soltani M, Teeley A, Jensen MP. Virtual reality hypnosis in the treatment of chronic neuropathic pain: a case report. Int J Clin Exp Hypn 2008 Oct;56(4):451-462 [FREE Full text] [doi: 10.1080/00207140802255534] [Medline: 18726807]
15. Patterson DR, Jensen MP, Wiechman SA, Sharar SR. Virtual reality hypnosis for pain associated with recovery from physical trauma. Int J Clin Exp Hypn 2010 Jul;58(3):288-300 [FREE Full text] [doi: 10.1080/00207141003760595] [Medline: 20509069]
16. Rousseaux F, Faymonville M, Nyssen A, Dardenne N, Ledoux D, Massion PB, et al. Can hypnosis and virtual reality reduce anxiety, pain and fatigue among patients who undergo cardiac surgery: a randomised controlled trial. Trials 2020 Apr 15;21(1):330 [FREE Full text] [doi: 10.1186/s13063-020-4222-6] [Medline: 32293517]
17. Thompson T, Steffert T, Steed A, Gruzelier J. A randomized controlled trial of the effects of hypnosis with 3-D virtual reality animation on tiredness, mood, and salivary cortisol. Int J Clin Exp Hypn 2011 Jan;59(1):122-142. [doi: 10.1080/00207144.2011.522917] [Medline: 21104488]
18. Barber TX, Wilson SC. The Barber Suggestibility Scale and the Creative Imagination Scale: experimental and clinical applications. Am J Clin Hypn 1978;21(2-3):84-108. [doi: 10.1080/00029157.1978.10403966] [Medline: 747170]
19. Spielberger C, Gorsuch R, Lushene R, Vagg P, Jacobs G. Manual for the State-Trait Anxiety Inventory. Palo Alto, CA: Consulting Psychologists Press; 1983.
20. Guétin S, Giniès P, Siou DKA, Picot M, Pommié C, Guldner E, et al. The effects of music intervention in the management of chronic pain: a single-blind, randomized, controlled trial. Clin J Pain 2012 May;28(4):329-337. [doi: 10.1097/AJP.0b013e31822be973] [Medline: 22001666]
21. Brunet D, Murray MM, Michel CM. Spatiotemporal analysis of multichannel EEG: CARTOOL. Comput Intell Neurosci 2011;2011:813870. [doi: 10.1155/2011/813870] [Medline: 21253358]
22. Logier R, Jeanne M, De Jonckheere J, Dassonneville A, Delecroix M, Tavernier B. PhysioDoloris: a monitoring device for analgesia / nociception balance evaluation using heart rate variability analysis. Annu Int Conf IEEE Eng Med Biol Soc 2010;2010:1194-1197. [doi: 10.1109/IEMBS.2010.5625971] [Medline: 21095676]
23. Makowski D, Pham T, Lau Z, Brammer J, Lespinasse F, Pham H, et al. NeuroKit2: A Python toolbox for neurophysiological signal processing. Behav Res Methods 2021 Aug;53(4):1689-1696. [doi: 10.3758/s13428-020-01516-y] [Medline: 33528817]

24. Colloca L, Raghuraman N, Wang Y, Akintola T, Brawn-Cinani B, Colloca G, et al. Virtual reality: physiological and behavioral mechanisms to increase individual pain tolerance limits. Pain 2020 Sep 01;161(9):2010-2021 [FREE Full text] [doi: 10.1097/j.pain.0000000000001900] [Medline: 32345915]
25. Patterson DR, Hoffman HG, Palacios AG, Jensen MJ. Analgesic effects of posthypnotic suggestions and virtual reality distraction on thermal pain. J Abnorm Psychol 2006 Nov;115(4):834-841. [doi: 10.1037/0021-843X.115.4.834] [Medline: 17100541]
26. Garcia-Larrea L, Frot M, Valeriani M. Brain generators of laser-evoked potentials: from dipoles to functional significance. Neurophysiol Clin 2003 Dec;33(6):279-292. [doi: 10.1016/j.neucli.2003.10.008] [Medline: 14678842]
27. García-Larrea L, Peyron R, Laurent B, Mauguière F. Association and dissociation between laser-evoked potentials and pain perception. Neuroreport 1997 Dec 01;8(17):3785-3789. [doi: 10.1097/00001756-199712010-00026] [Medline: 9427371]
28. Craig AD. A new view of pain as a homeostatic emotion. Trends Neurosci 2003 Jun;26(6):303-307. [doi: 10.1016/s0166-2236(03)00123-1] [Medline: 12798599]
29. Schweinhardt P, Bushnell MC. Pain imaging in health and disease--how far have we come? J Clin Invest 2010 Nov;120(11):3788-3797. [doi: 10.1172/JCI43498] [Medline: 21041961]
30. Lier EJ, Oosterman JM, Assmann R, de Vries M, van Goor H. The effect of virtual reality on evoked potentials following painful electrical stimuli and subjective pain. Sci Rep 2020 Jun 03;10(1):9067. [doi: 10.1038/s41598-020-66035-4] [Medline: 32494060]
31. Elkins GR, Barabasz AF, Council JR, Spiegel D. Advancing research and practice: the revised APA Division 30 definition of hypnosis. Int J Clin Exp Hypn 2015;63(1):1-9. [doi: 10.1080/00207144.2014.961870] [Medline: 25365125]
32. Wickramasuriya DS, Faghih RT. A mixed filter algorithm for sympathetic arousal tracking from skin conductance and heart rate measurements in Pavlovian fear conditioning. PLoS One 2020;15(4):e0231659 [FREE Full text] [doi: 10.1371/journal.pone.0231659] [Medline: 32324756]
33. Jafari H, Courtois I, Van den Bergh O, Vlaeyen JWS, Van Diest I. Pain and respiration: a systematic review. Pain 2017 Jun;158(6):995-1006. [doi: 10.1097/j.pain.0000000000000865] [Medline: 28240995]
34. Patterson DR, Wiechman SA, Jensen M, Sharar SR. Hypnosis delivered through immersive virtual reality for burn pain: a clinical case series. Int J Clin Exp Hypn 2006 Apr;54(2):130-142. [doi: 10.1080/00207140500528182] [Medline: 16581687]
35. El Tumi H, Johnson MI, Dantas PBF, Maynard MJ, Tashani OA. Age-related changes in pain sensitivity in healthy humans: A systematic review with meta-analysis. Eur J Pain 2017 Jul;21(6):955-964. [doi: 10.1002/ejp.1011] [Medline: 28230292]
36. Lanitis S, Mimigianni C, Raptis D, Sourtse G, Sgourakis G, Karaliotas C. The impact of educational status on the postoperative perception of pain. Korean J Pain 2015 Oct;28(4):265-274 [FREE Full text] [doi: 10.3344/kjp.2015.28.4.265] [Medline: 26495081]
37. Yezierski RP. The effects of age on pain sensitivity: preclinical studies. Pain Med 2012 Apr;13(Suppl 2):S27-S36 [FREE Full text] [doi: 10.1111/j.1526-4637.2011.01311.x] [Medline: 22497745]
38. Garcia LM, Darnall BD, Krishnamurthy P, Mackey IG, Sackman J, Louis RG, et al. Correction: Self-administered behavioral skills-based at-home virtual reality therapy for chronic low back pain: protocol for a randomized controlled trial. JMIR Res Protoc 2021 Feb 12;10(2):e27652 [FREE Full text] [doi: 10.2196/27652] [Medline: 33577465]
39. Garcia LM, Birckhead BJ, Krishnamurthy P, Mackey I, Sackman J, Salmasi V, et al. Three-month follow-up results of a double-blind, randomized placebo-controlled trial of 8-week self-administered at-home behavioral skills-based virtual reality (VR) for chronic low back pain. J Pain 2022 May;23(5):822-840 [FREE Full text] [doi: 10.1016/j.jpain.2021.12.002] [Medline: 34902548]

Abbreviations

ANI: analgesia nociception index
ANR: Agence Nationale de la Recherche (French national research agency)
ANOVA: analysis of variance
CHEP: contact heat evoked potential
ECG: electrocardiogram
EEG: electroencephalogram
pNN50: percentage of successive R-R intervals that differ by more than 50 milliseconds
RMSSD: root mean square of successive R-R interval differences
SCR: skin conductance response
STAI: State-Trait Anxiety Inventory
VR: virtual reality
VRH: virtual reality hypnosis

Edited by R Kukafka; submitted 15.09.21; peer-reviewed by S Persky, M O'Brien; comments to author 29.12.21; revised version received 11.02.22; accepted 13.06.22; published 29.07.22.

Please cite as:
Terzulli C, Melchior M, Goffin L, Faisan S, Gianesini C, Graff D, Dufour A, Laroche E, Chauvin C, Poisbeau P
Effect of Virtual Reality Hypnosis on Pain Threshold and Neurophysiological and Autonomic Biomarkers in Healthy Volunteers: Prospective Randomized Crossover Study
J Med Internet Res 2022;24(7):e33255
URL: https://www.jmir.org/2022/7/e33255
doi:10.2196/33255
PMID:35904872

Original Paper

Wisdom of the Experts Versus Opinions of the Crowd in Hospital Quality Ratings: Analysis of Hospital Compare Star Ratings and Google Star Ratings

Hari Ramasubramanian[1*], PhD; Satish Joshi[2*], PhD; Ranjani Krishnan[3*], PhD

[1]Accounting Department, Frankfurt School of Finance and Management, Frankfurt am Main, Germany

[2]College of Agriculture & Natural Resources, Department of Agricultural, Food, and Resource Economics, Michigan State University, East Lansing, MI, United States

[3]Accounting and Information Systems, Broad College of Business, Michigan State University, East Lansing, MI, United States

[*]all authors contributed equally

Corresponding Author:
Hari Ramasubramanian, PhD
Accounting Department
Frankfurt School of Finance and Management
32-34 Adickesallee
Frankfurt am Main, 60320
Germany
Phone: 49 69154008823
Email: h.ramasubramanian@fs.de

Abstract

Background: Popular web-based portals provide free and convenient access to user-generated hospital quality reviews. The Centers for Medicare & Medicaid Services (CMS) also publishes Hospital Compare Star Ratings (HCSR), a comprehensive expert rating of US hospital quality that aggregates multiple measures of quality. CMS revised the HCSR methods in 2021. It is important to analyze the degree to which web-based ratings reflect expert measures of hospital quality because easily accessible, crowdsourced hospital ratings influence consumers' hospital choices.

Objective: This study aims to assess the association between web-based, Google hospital quality ratings that reflect the opinions of the crowd and HCSR representing the wisdom of the experts, as well as the changes in these associations following the 2021 revision of the CMS rating system.

Methods: We extracted Google star ratings using the Application Programming Interface in June 2020. The HCSR data of April 2020 (before the revision of HCSR methodology) and April 2021 (after the revision of HCSR methodology) were obtained from the CMS Hospital Compare website. We also extracted scores for the individual components of hospital quality for each of the hospitals in our sample using the code provided by Hospital Compare. Fractional response models were used to estimate the association between Google star ratings and HCSR as well as individual components of quality (n=2619).

Results: The Google star ratings are statistically associated with HCSR (P<.001) after controlling for hospital-level effects; however, they are not associated with clinical components of HCSR that require medical expertise for evaluation such as safety of care (P=.30) or readmission (P=.52). The revised CMS rating system ameliorates previous partial inconsistencies in the association between Google star ratings and quality component scores of HCSR.

Conclusions: Crowdsourced Google star hospital ratings are informative regarding expert CMS overall hospital quality ratings and individual quality components that are easier for patients to evaluate. Improvements in hospital quality metrics that require expertise to assess, such as safety of care and readmission, may not lead to improved Google star ratings. Hospitals can benefit from using crowdsourced ratings as timely and easily available indicators of their quality performance while recognizing their limitations and biases.

(J Med Internet Res 2022;24(7):e34030) doi:10.2196/34030

KEYWORDS

hospital quality; web-based rating; online ratings; Hospital Compare; star ratings

Introduction

In recent years, crowdsourced, web-based information aggregated by social media platforms, service providers, and government agencies has become a popular source for consumers, organizations, and governments. If used effectively, crowdsourcing can gather timely feedback, improve efficiency and supervision, reduce response times for critical actions, and increase customer engagement and satisfaction [1]. Although crowdsourced ratings are widely used, including in health care settings, concerns arise about crowdsourced assessments due to the following reasons: (1) nonrepresentative nature of the sampling process, (2) inadequacies in expertise required to assess the quality of the goods or services provided, (3) the limited nature of experiences, and (4) the potential for manipulation [1-4]. Especially for complex services such as medical treatment, consumers may not have the expertise to evaluate the quality of care even after the completion of the consumption cycle, and crowdsourced ratings may have little or no correspondence with true quality, due to cognitive limitations and expertise gap [5].

In this analysis, we explored the association between a popular crowdsourced, web-based hospital quality rating (Google star) and expert rating of hospital quality (HCSR) published by the Centers for Medicare & Medicaid Services (CMS) [6]. Understanding the association between crowdsourced opinions and expert hospital quality ratings is important because web-based ratings appear to have substantial influence on patients' health care choices [7]. For example, 65% of surveyed patients in 2020 used web-based reviews to evaluate doctors, and Google was the most visited review site [8,9]. In comparison, only 22% of the surveyed patients were even aware of CMS's expert ratings [10].

While the Hospital Compare website provides information on more than 100 quality measures for more than 4000 US hospitals, HCSR aggregates multiple quality measures into a single overall star rating to increase user convenience. CMS revised the HCSR methodology in 2021. Previously, 57 measures of hospital quality were assigned to the 7 clusters of hospital quality, namely, "patient experience," "mortality," "readmission," "safety of care," "efficient use of medical imaging," "timeliness of care," and "effectiveness of care," and a hospital-specific score was derived for each of these groups using a latent variable statistical model. Hospital summary score was computed as a weighted sum of its group scores, and hospitals were assigned to the star categories based on a clustering algorithm [11-13]. The revised HCSR method reduced the number of quality measures to 48 and quality clusters to 5 by combining the separate measures for "timeliness of care", "effectiveness of care", and "efficient use of medical imaging." The latent variable model was replaced by a simple average of individual measures making the scores transparent, understandable, and predictable [12].

We empirically analyzed the associations between Google star and HCSR before and after the revisions. We also analyzed the associations with HCSR quality component scores. The analyses reveal that Google star ratings are statistically associated with HCSR (P<.001) after controlling for hospital-level effects but are not associated with clinical components of HCSR that require medical expertise for evaluation, such as safety of care (P=.30) or readmission (P=.52), after the 2021 HCSR revision.

While a number of prior studies have analyzed the relationship between crowdsourced hospital ratings (eg, Yelp and Facebook) and expert ratings [14-17], our study contributes to the literature by examining the association between the more popular Google star ratings and expert HCSR ratings before and after the recent revision of HCSR methodology. Unlike the previous research, which has analyzed associations with some selected clinical quality indicators [18,19], we analyzed the associations with all the individual HCSR quality cluster scores. Additionally, we analyzed these associations by employing a more refined statistical technique of fractional response modeling that accounts for boundedness and the nonlinear nature of the star ratings while controlling for other covariates [20].

Methods

Data Source

We used the National Bureau of Economic Research's Healthcare Cost Report Information System to generate a data set of 4615 US hospitals and their characteristics such as the size (number of beds), type in terms of location (rural or urban), and status, that is, for-profit or nonprofit and teaching or nonteaching. We collected HCSR data from the CMS Hospital Compare website in April 2020 before the revision of HCSR, and again in April 2021 after the revision. We also extracted scores for the individual components of hospital quality for each of the hospitals in our sample using the code provided by CMS. In June 2020, we extracted Google star ratings for all available US hospitals using the Application Programming Interface, which is a standard interface for collecting data from web-based portals. Google Places Application Programming Interface provides Google's Star ratings, which are the cumulative average of consumer ratings of each hospital, ranging between 1 and 5. To assure quality, Google removes all anonymous reviews, requires a valid associated email address, and does not allow more than one review per business from a particular email [21]. Because the 2020 HCSR draw on hospital quality data covering the period 2015-2018, and the 2021 HCSR use data for the period 2016-2019, we used the Google star ratings extracted in June 2020, which are cumulative ratings and useful for analyzing associations both before and after the HCSR revision. This facilitated the consistent analysis of these associations, especially in view of the potential impact of the COVID-19 pandemic on subsequent Google star hospital ratings. A total of 2963 US hospitals had received HCSR and all the component scores in both 2020 and 2021 reporting cycles, and the 46 (1.6%) hospitals that did not have Google star ratings were excluded from analyses. To ensure that Google star ratings were representative, and following a similar approach used in previous studies [14], we excluded 298 (10.1%) hospitals, which had less than 10 individual Google ratings. This resulted in the final sample of 2619 US hospitals with an average of 179 individual Google ratings per hospital. However, including hospitals with

less than 10 individual Google ratings in our estimations resulted in similar conclusions.

Research Design

We theorize that changes in true hospital quality influence both HCSR expert ratings and crowdsourced, web-based Google star ratings, and as a result, HCSR and Google star ratings will be correlated. However, hospital-level characteristics such as size and type are also likely to affect consumer perceptions of hospital quality and therefore web-based ratings. Unlike HCSR that factors in hospital-level drivers of quality or risk, Google star ratings are not adjusted for hospital characteristics. Therefore, it is important to control for these while assessing associations between Google star and HCSR. Further, improvements in health quality component scores improve hospital quality and, by design, their aggregate HCSR; we thus hypothesize that crowdsourced Google star ratings will also be associated with HCSR quality component scores. While it would be ideal to explore the association of HCSR component scores with similar quality component scores in Google star ratings (which are not available) or some proxies for quality clusters developed from detailed textual analyses of accompanying comments, we posit that associations between aggregate Google star ratings and HCSR component scores will be informative approximations.

We estimated the following models to examine the association between Google hospital ratings, HCSR, and individual HCSR component scores, after controlling for hospital size and hospital type effects on web-based ratings.

Where  is the association between Google star ratings and HCSR,  is a vector of control variables for hospital *i*, which includes size (logarithm of the number of beds), and hospital type (for-profit, not-for-profit, or government hospital; teaching or nonteaching hospital) and their interaction terms. $HCSRcomponent_j$ is the HCSR component score for each cluster of hospital quality. Our dependent variable, Google star ratings, is a continuous variable bounded between 1 and 5, and linear estimation methods can produce predicted values outside these bounds. Therefore, we scaled the Google star ratings to be bounded between 0 and 1 and employ a fractional response model (FRM) to estimate these equations [20]. The FRM is an extension of the generalized linear model that accounts for the boundedness of the dependent variable from both above and below, predicts response values within the interval limits of the dependent variable, and captures the nonlinearity of the data, thereby yielding a better fit compared to linear estimation models. Furthermore, the FRM does not require special data transformations at the corners; it also permits a robust, consistent, and relatively efficient estimation of the conditional expectation of the dependent variable given the predictors [20].

Results

Figure 1 shows the distributions of "HCSR2020," "HCSR2021" (the Hospital Compare Star ratings provided by CMS in April 2020 and April 2021, respectively) and "Google star" ratings. The means of HCSR2020, HCSR2021, and Google star ratings are similar with values of 3.159, 3.236, and 3.040, respectively. However, the distributions of HCSR are relatively wider when compared to Google star ratings, as evidenced by standard deviations of 1.133, 1.114, and 0.557, respectively. In Figure 1, Google star ratings have been rounded up or down to the nearest integer in the bar graph. Accordingly, HCSR2020, HCSR2021, and Google star ratings in the graph can take integer values from 1 and 5.

Simple correlation analyses show statistically significant (P<.001) Pearson correlation coefficients of 0.234 between Google star ratings and HCSR2020, and 0.226 between Google star and HCSR2021 ratings. Spearman correlation coefficients are 0.242 between Google star and HCSR2020, and 0.224 between Google star and HCSR2021, and both are significant (P<.001).

Since the above results do not control for variations in hospital characteristics, we estimate the parameters of Equation (1) employing FRM regression techniques with scaled Google star ratings as a dependent variable; we also estimate HCSR as well as hospital size and type as explanatory variables. Table 1 reports the estimation results for the relationship between Google star ratings and Hospital Compare Star Ratings (HCSR). The dependent variable is Google star ratings scaled by 5 to be between 0 and 1. The results are estimated using fractional response probit regression with standard errors robust to distributional misspecification. The coefficients on HCSR2020 and HCSR2021 are 0.076 and 0.070 and statistically significant (P<.001). The statistically significant coefficient estimates on HCSR confirm that Google star ratings have informative value even after controlling for hospital factors. Similarly, significant coefficients on hospital size and type variables (eg, for-profit, teaching, rural, and their interactions) support our hypothesis that hospital characteristics influence consumer perceptions of hospital quality.

Table 2, column 1 shows FRM estimates of the association between HCSR2020 and its individual components, while column 2 shows FRM estimation results (Equation 2) with Google star ratings as the dependent variable and HCSR2020 component scores of hospital quality as explanatory variables, with controls for hospital size and type. Columns 3 and 4 show similar estimation results using the component scores for HCSR2021. In all the columns, dependent variable is scaled by dividing by 5 to be between 0 and 1. The results are estimated using fractional response probit regression with standard errors robust to distributional misspecification.

As can be seen from columns 1 and 3, both HCSR2020 and HCSR2021 are positively associated with all the component scores (P<.001), which is unsurprising since overall HCSR is assigned based on the weighted sum of individual scores. The results in column 2 show that Google star ratings had statistically significant positive association with patient experience (P<.001),

mortality (*P*<.001), and effectiveness of care (*P*<.001) components of HCSR2020. Google star ratings were negatively associated with the 2020 quality scores for readmission (*P*=.01) and efficient use of medical imaging (*P*=.06), indicating that hospitals considered by experts as high performing on these 2 quality dimensions were likely to be perceived as lower quality by the public. In comparison, the component scores for patient experience (*P*<.001), mortality (*P*<.001), and the new combined timeliness and effectiveness of care (*P*<.001) of the revised HCSR2021 are all positively associated with Google star ratings (as shown in Table 2, column 4). Moreover, the association with readmission, which was previously negative, has become statistically insignificant (*P*=.52). Variations in the HCSR component score on safety of care, either before or after the revisions, do not significantly affect Google star ratings (*P*=.49 and *P*=.30).

The coefficient estimates in Table 2 are difficult to interpret directly because the FRM estimation technique is nonlinear, the scaling of Google star ratings is between 0 and 1, and there are adjustments needed to account for different weightings of individual components in the final HCSR. Therefore, we calculated the average marginal effects of individual quality components and their 95% confidence intervals on Google star ratings and HCSR2021 (Figure 2; data source: Table 2, columns 3 and 4). Average marginal effect is the average of the marginal effects computed for every observation in the sample [22].

The average marginal effect of patient experience in the Google star ratings is 1.143 (*P*<.001), that is, 1 unit increase in patient experience scores of HCSR is expected to increase the Google star ratings by 1.143, on average. Similarly, the average marginal effect of mortality is 0.195 (*P*<.001), readmission is –0.035 (*P*=.52), safety of care is 0.051 (*P*=.30), and timeliness and effectiveness of care is 0.489 (*P*<.001). In comparison, the average marginal effects of all dimensions of quality on HCSR are statistically significant (*P*<.001). The average marginal effects of patient experience, mortality, readmission, safety of care, and timeliness and effectiveness of care are 1.942, 2.236, 2.260, 1.988, and 2.249, respectively.

Figure 1. Graph showing HCSR2020, HCSR2021, and Google Star ratings for the sample of 2619 US hospitals. HCSR: Hospital Compare Star Ratings.

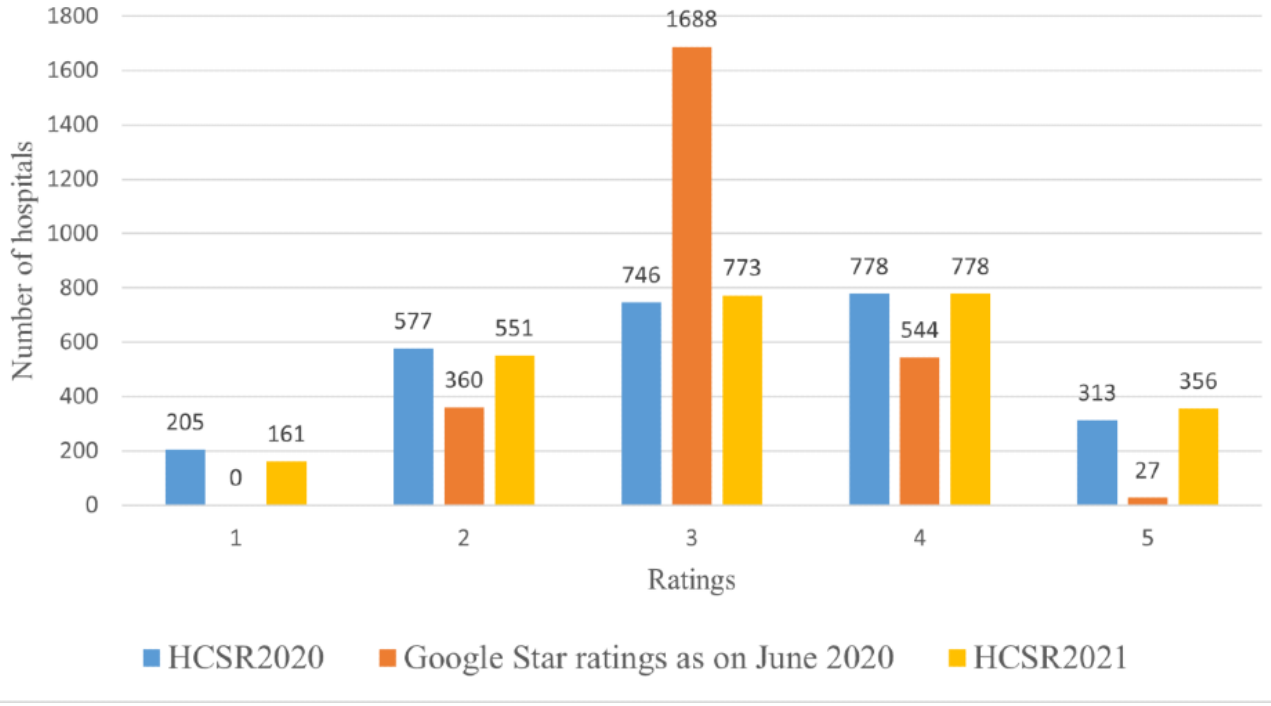


Table 1. Fractional response probit regression of Google star ratings on HCSR[a]2020 and HCSR2021.

Variable	Google star rating[b]	*P* value	Google star rating[c]	*P* value
HCSR	0.076	<.001	0.070	<.001
Log beds	0.022	.005	0.009	.23
For-profit	0.165	<.001	0.165	<.001
Teaching	0.069	<.001	0.076	<.001
For-profit × teaching	0.007	.85	0.010	.79
Rural	–0.004	.81	0.009	.59
Rural × for-profit	–0.143	<.001	–0.155	<.001
Rural × teaching	–0.058	.04	–0.073	.009
Rural × for-profit × teaching	–0.037	.62	–0.045	.56
Constant	–0.015	.44	–0.005	.81
Observations	2619	N/A[d]	2619	N/A
Pseudo *R*-squared	0.005	N/A	0.004	N/A

[a]HCSR: Hospital Compare Star Ratings.

[b]This column reports estimation results using Google star ratings as a dependent variable with HCSR2020 and hospital controls.

[c]This column reports estimation results using Google star ratings as a dependent variable with HCSR2021 and hospital controls.

[d]N/A: not applicable.

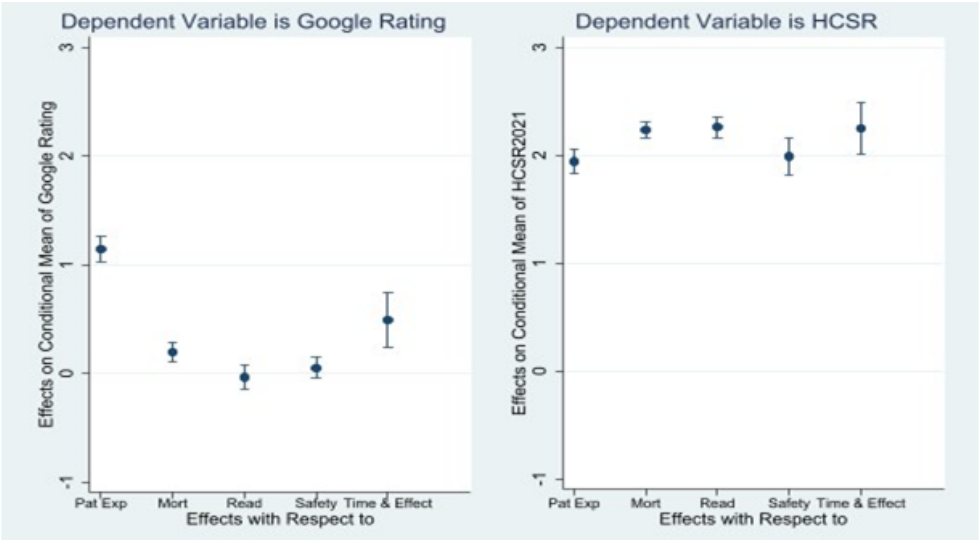
Table 2. Fractional response probit regression of Google star ratings and HCSR[a] on individual component scores before and after the change in methodology of computing HCSR.

Variables	(1) HCSR2020	*P* value	(2) Google star ratings	*P* value	(3) HCSR2021	*P* value	(4) Google star ratings	*P* value
Patient experience	1.376	<.001	0.640	<.001	1.249	<.001	0.600	<.001
Mortality	1.377	<.001	0.112	<.001	1.438	<.001	0.102	<.001
Readmission	1.378	<.001	–0.058	.01	1.454	<.001	–0.018	.52
Safety of care	1.272	<.001	0.016	.49	1.279	<.001	0.027	.30
Efficient use of medical imaging	1.370	<.001	–0.276	.06	__[b]	—	—	—
Timeliness of care	0.924	<.001	0.336	.07	—	—	—	—
Effectiveness of care	1.720	<.001	1.241	<.001	—	—	—	—
Timeliness and effectiveness of care	—	—	—	—	1.446	<.001	0.257	<.001
Log beds	0.005	.52	0.046	<.001	0.091	<.001	0.055	<.001
For-profit	–0.015	.39	0.201	<.001	–0.054	.007	0.199	<.001
Teaching	0.002	.90	0.067	<.001	–0.044	.005	0.060	<.001
For-profit × teaching	0.019	.49	–0.026	.45	0.009	.76	0.001	.98
Rural	–0.008	.55	–0.022	.14	–0.110	<.001	–0.029	.05
Rural × for-profit	–0.009	.74	–0.155	<.001	0.027	.47	–0.148	<.001
Rural × teaching	–0.001	.97	–0.056	.03	0.108	<.001	–0.042	.12
Rural × for-profit × teaching	0.063	.39	0.036	.60	0.036	.58	–0.005	.94
Constant	0.480	<.001	0.241	<.001	0.594	<.001	0.247	<.001
Observations	2619	—	2619	—	2619	—	2619	—
Pseudo *R*-squared	0.166	—	0.009	—	0.154	—	0.008	—

[a]HCSR: Hospital Compare Star Ratings.

[b]Not applicable.

Figure 2. Average marginal effects of individual quality components of Hospital Compare Star Ratings (HCSR) issued by Centers for Medicare & Medicaid Services in April 2021 (HCSR2021) on Google star ratings and HCSR2021. Pat Exp: patient experience; Mort: mortality; Read: readmission; Safety: safety of care; Time & Effect: timeliness & effectiveness of care.



Discussion

Principal Findings

Our analysis adds to the stream of prior studies analyzing the associations between social media ratings and HCSR or Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) ratings and other selected measures of clinical quality [3]. A recent systematic literature review identified 32 peer reviewed studies (25 using US data) published before 2020, which quantitatively examined relations between web-based ratings and health care outcomes [18]. These include studies by Bardach et al [14] and Ranard et al [15], which find a significant correlation between Yelp hospital ratings and HCAHPS survey scores of patient experience, and studies by Campbell and Li [16] and Huppetz and Otto [23], which find a positive association between Facebook star ratings and HCAHPS scores. Hawkins et al [24] find a positive association between the percentage of patients with top HCAHPS scores (>9) and the sentiment score derived from text analysis of Twitter posts. Subsequent studies also report similar findings about associations between web-based ratings (Yelp and Facebook) and HCSR and HCAHPS scores [17,19].

We refined those analyses by examining the associations between Google star ratings and HCSR and its individual quality components before and after the 2021 revision of HCSR methodology, and by using a more appropriate statistical estimation technique. We used Google star ratings because many studies report that patients find providers by first "googling" their symptom and then "googling" providers that care for conditions associated with their symptoms, suggesting that Google is increasingly used by consumers to search for and access care [7-9]. However, only one prior study has analyzed Google star ratings [19].

Our correlation analysis yields statistically significant positive correlations between crowdsourced Google star ratings and HCSR, confirming that Google star ratings provide directional information that is consistent with the expert HCSR. However, the narrower distribution of Google star ratings indicates that patients tend to avoid extreme ratings in web-based reviews, resulting in relatively lower Google star ratings for highest quality hospitals and vice versa. To the extent that web-based ratings influence hospital choices at the margin, the lower variance of Google star ratings may result in a relative underchoosing of highest-quality hospitals and overchoosing of lower-quality hospitals. FRM estimation with hospital type and size controls supports our conjecture that hospital-level characteristics such as size and type affect consumer perceptions of hospital quality and therefore web-based ratings. Estimation results in Table 1 (columns 1 and 2) show statistically significant positive coefficients on for-profit and teaching (P<.001), indicating that these hospitals receive higher Google star ratings compared to the baseline of urban, nonprofit, and nonteaching hospitals, even after controlling for the differences in hospital quality as measured by HCSR. Similarly, the negative coefficients on interaction terms rural × for-profit (P<.001) and rural × teaching (P=.04 and P=.009) indicate that these hospitals receive relatively lower ratings. The statistically significant positive coefficient on log beds (P=.005) in Table 1 (column 1) suggests that larger hospitals received higher Google star ratings even after controlling for the differences in hospital quality and type. However, with the revised HCSR, this bias in Google star ratings favoring large hospitals appears to have been attenuated (P=.23; Table 1, column 2).

Consistent with the findings in most prior studies that web-based ratings are associated with patient experience HCAHPS quality scores [14-16,23,24], our FRM estimates of the relationship between HCSR and its individual components in Table 2 show a positive association between Google star ratings and HCSR patient experience scores. Google star ratings also exhibit positive associations with component scores for mortality, timeliness of care, and effectiveness of care in HCSR2020 and the combined score for timeliness and effectiveness of care in HCSR2021, which suggests that patients are able to evaluate other quality dimensions at least partially. For example, the average time spent in the emergency room, an underlying measure of timeliness of care, and the percent of newborns whose deliveries were scheduled too early when medically not necessary—a measure used in assessing effectiveness of care, can be assessed by patients who usually do not possess medical expertise. The HCSR2020 quality scores for readmission and efficient use of medical imaging were negatively associated with Google star ratings, potentially leading to suboptimal choices by consumers of hospitals based on their perceptions regarding readmission and efficient use of medical imaging. The updated HCSR2021 does not provide a separate score for medical imaging, and the prior negative association between Google star ratings and readmission has become statistically insignificant, suggesting that the revised HCSR methods may partially ameliorate suboptimal hospital choices based on Google star ratings from previous rating methods. All together, these results suggest that crowdsourced Google star ratings have directional information value that is consistent with expert ratings on select dimensions of hospital quality.

A comparison of the marginal effects of individual components of quality on Google star ratings and HCSR (Figure 2) reveals a consistent pattern of underweighting of medical quality by patients who provide Google star ratings. As mentioned before, the overall variance of Google star ratings is less than the variance in HCSR because patients appear to avoid extreme ratings, which in turn explains the relatively lower sensitivity of Google star ratings to improvements in the component scores compared to HCSR. Moreover, because patients are not exposed to or are unable to assess the quality measures underlying readmission and safety of care, Google star ratings appear to be unaffected by the improvements in these quality components. In other words, hospitals need to make significant improvements in patient experience, mortality, as well as timeliness and effectiveness of care to improve their Google star ratings but should not expect improvements in safety of care and readmission to result in improved Google star ratings.

Limitations

Our observations are subject to several caveats. Obviously, consumer choice of hospitals is a very complex decision guided not simply by summary ratings of hospital quality but by

medical condition, financial situation, urgency, insurance status, network access, hospital location, and provider preferences. While our study does not analyze how web-based reviews affect actual consumer hospital choices, we draw on other studies [7,8,21], which show that hospital choices are influenced by web-based ratings and conjecture their marginal effects on hospital choices. Our analysis is based on a single source of web-based ratings and expert ratings in the United States and may not be generalizable across other countries and rating sources. We analyzed the associations between aggregate Google star ratings and HCSR and its component scores; however, we overlooked the potentially rich data contained in detailed web-based review comments. Additionally, Google star ratings are cumulative averages, while HCSR is based on the previous 3 years of data, which may result in mismatched data and lags in responsiveness. Greater transparency from Google about its algorithms would enable better comparability. Even the revised HCSR method has been criticized for the following reasons: using relative performance–based ratings instead of predefined standards of performance, lack of true peer group comparisons and risk adjustment, inadequate audit or verification of data, and dependence on advisory technical expert panels instead of rigorous peer review [13]. Future revisions to HCSR methods to address these issues will necessitate an updated analyses of associations of web-based ratings with HCSR.

Implications for Research and Practice

Consistent with previous findings, our analysis shows that aggregate Google star ratings provide directional information consistent with the expert HCSR and are also associated with selected HCSR quality component scores related to patient experience and other components that patients can partially assess (eg, mortality as well as timeliness and effectiveness of care). While our results suggest even the aggregate web-based ratings are informative, future research employing natural language processing and sentiment analyses techniques on detailed web-based review comments and assessing their associations with quality measures in HCSR can generate more nuanced insights. Research is needed to analyze the causes of the observed divergence between web-based ratings and specific components of expert ratings. Hospital rating agencies such as CMS need to launch education efforts to address these knowledge gaps and to increase consumers' use of expert ratings. Hospitals can benefit from using crowdsourced ratings as timely, accessible, and dynamic indicators of their quality performance, while keeping in mind their sensitivities and biases. Because of their universal and timely availability, crowdsourced data along with personal health applications, such as Google Health and Microsoft HealthVault, and electronic medical records will continue to evolve as important components of the larger "digital transformation" of health care [25]. Research to assess the interactive effects of these on decision-making by patients as well as health care providers is crucial.

Acknowledgments

The authors received no financial support for the research, authorship, and publication of this paper. The authors acknowledge the research assistance provided by Hetal Adhia, Harshad Badyani, Priya Jindal, and Dheeraj Kasar.

Conflicts of Interest

None declared.

References

1. Kleefstra SM, Zandbelt LC, Borghans I, de Haes HJCJM, Kool RB. Investigating the Potential Contribution of Patient Rating Sites to Hospital Supervision: Exploratory Results From an Interview Study in the Netherlands. J Med Internet Res 2016 Jul 20;18(7):e201 [FREE Full text] [doi: 10.2196/jmir.5552] [Medline: 27439392]
2. Weil AG, Bojanowski MW, Jamart J, Gustin T, Lévêque M. Evaluation of the quality of information on the Internet available to patients undergoing cervical spine surgery. World Neurosurg 2014;82(1-2):e31-e39. [doi: 10.1016/j.wneu.2012.11.003] [Medline: 23142585]
3. Verhoef LM, Van de Belt TH, Engelen LJLPG, Schoonhoven L, Kool RB. Social media and rating sites as tools to understanding quality of care: a scoping review. J Med Internet Res 2014 Feb 20;16(2):e56 [FREE Full text] [doi: 10.2196/jmir.3024] [Medline: 24566844]
4. Kool RB, Kleefstra SM, Borghans I, Atsma F, van de Belt TH. Influence of Intensified Supervision by Health Care Inspectorates on Online Patient Ratings of Hospitals: A Multilevel Study of More Than 43,000 Online Ratings. J Med Internet Res 2016 Jul 15;18(7):e198 [FREE Full text] [doi: 10.2196/jmir.5884] [Medline: 27421302]
5. Ekelund RB, Mixon FG, Ressler RW. Advertising and information. Journal of Economic Studies 1995 Apr 01;22(2):33-43 [FREE Full text] [doi: 10.1108/01443589510086970]
6. Hospital Compare. Centers for Medicare & Medicaid Services. 2019. URL: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalCompare [accessed 2019-12-01]
7. How first impressions online affect patient acquisition. Becker's Hospital Review. 2017. URL: https://www.beckershospitalreview.com/hospital-management-administration/how-first-impressions-online-affect-patient-acquisition-and-hospital-revenue.html [accessed 2019-09-28]

8. Hedges L, Couey C. How patients use online reviews. Software Advice. 2020. URL: https://www.softwareadvice.com/resources/how-patients-use-online-reviews/#back [accessed 2021-09-28]
9. Introducing the 2021 Healthcare Reputation Report. Reputation.com, Inc. URL: https://reputation.com/resources/articles/introducing-the-2021-healthcare-reputation-report/ [accessed 2021-07-03]
10. One in five patients use CMS star ratings for healthcare decisions. Patient Engagement Hit. URL: https://patientengagementhit.com/news/one-in-five-patients-use-cms-star-ratings-for-healthcare-decisions [accessed 2021-09-28]
11. DeLancey JO, Softcheck J, Chung JW, Barnard C, Dahlke AR, Bilimoria KY. JAMA 2017 May 16;317(19):2015-2017 [FREE Full text] [doi: 10.1001/jama.2017.3148] [Medline: 28510670]
12. Overall hospital quality star rating on hospital compare public input request. Centers for Medicare & Medicaid Services. 2019. URL: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/Downloads/Overall-Hospital-Quality-Star-Rating-on-Hospital-Compare-Public-Input-Period.pdf [accessed 2019-09-28]
13. Bilimoria KY, Barnard C. An Evolving Hospital Quality Star Rating System From CMS: Aligning the Stars. JAMA 2021 Jun 01;325(21):2151-2152. [doi: 10.1001/jama.2021.6946] [Medline: 33999111]
14. Bardach NS, Asteria-Peñaloza R, Boscardin WJ, Dudley RA. The relationship between commercial website ratings and traditional hospital performance measures in the USA. BMJ Qual Saf 2013 Mar 23;22(3):194-202 [FREE Full text] [doi: 10.1136/bmjqs-2012-001360] [Medline: 23178860]
15. Ranard BL, Werner RM, Antanavicius T, Schwartz HA, Smith RJ, Meisel ZF, et al. Yelp Reviews Of Hospital Care Can Supplement And Inform Traditional Surveys Of The Patient Experience Of Care. Health Aff (Millwood) 2016 Apr;35(4):697-705 [FREE Full text] [doi: 10.1377/hlthaff.2015.1030] [Medline: 27044971]
16. Campbell L, Li Y. Are Facebook user ratings associated with hospital cost, quality and patient satisfaction? A cross-sectional analysis of hospitals in New York State. BMJ Qual Saf 2018 Feb;27(2):119-129. [doi: 10.1136/bmjqs-2016-006291] [Medline: 28600452]
17. Chakraborty S, Church EM. Social media hospital ratings and HCAHPS survey scores. J Health Organ Manag 2020 Feb 08;34(2):162-172 [FREE Full text] [doi: 10.1108/JHOM-08-2019-0234] [Medline: 32239874]
18. Placona AM, Rathert C. Are Online Patient Reviews Associated With Health Care Outcomes? A Systematic Review of the Literature. Med Care Res Rev 2022 Feb;79(1):3-16. [doi: 10.1177/10775587211014534] [Medline: 34027743]
19. Perez V, Freedman S. Do Crowdsourced Hospital Ratings Coincide with Hospital Compare Measures of Clinical and Nonclinical Quality? Health Serv Res 2018 Dec;53(6):4491-4506 [FREE Full text] [doi: 10.1111/1475-6773.13026] [Medline: 30084168]
20. Papke LE, Wooldridge JM. Econometric methods for fractional response variables with an application to 401(k) plan participation rates. J. Appl. Econ 1996 Nov;11(6):619-632. [doi: 10.1002/(sici)1099-1255(199611)11:6<619::aid-jae418>3.0.co;2-1]
21. What Google's move to drop anonymous reviews means for local brands. Mindstream Media Group. URL: https://mindstreammediagroup.com/what-googles-move-to-drop-anonymous-reviews-means-for-local-brands [accessed 2022-04-28]
22. Wooldridge J. Introductory Econometrics: A Modern Approach. Boston, USA: Cengage Learning; 2016.
23. Huppertz JW, Otto P. Predicting HCAHPS scores from hospitals' social media pages: A sentiment analysis. Health Care Manage Rev 2018;43(4):359-367. [doi: 10.1097/HMR.0000000000000154] [Medline: 28225448]
24. Hawkins JB, Brownstein JS, Tuli G, Runels T, Broecker K, Nsoesie EO, et al. Measuring patient-perceived quality of care in US hospitals using Twitter. BMJ Qual Saf 2016 Jun;25(6):404-413 [FREE Full text] [doi: 10.1136/bmjqs-2015-004309] [Medline: 26464518]
25. Kukafka R. Digital Health Consumers on the Road to the Future. J Med Internet Res 2019 Nov 21;21(11):e16359 [FREE Full text] [doi: 10.2196/16359] [Medline: 31750835]

Abbreviations

CMS: Center for Medicare & Medicaid Services
FRM: Fractional Response Model
HCAHPS: Hospital Consumer Assessment of Healthcare Providers and Systems
HCSR: Hospital Compare Star Ratings

Edited by R Kukafka; submitted 04.10.21; peer-reviewed by T Kool, M Richardson; comments to author 16.03.22; revised version received 08.05.22; accepted 13.06.22; published 26.07.22.

Please cite as:
Ramasubramanian H, Joshi S, Krishnan R
Wisdom of the Experts Versus Opinions of the Crowd in Hospital Quality Ratings: Analysis of Hospital Compare Star Ratings and Google Star Ratings
J Med Internet Res 2022;24(7):e34030
URL: https://www.jmir.org/2022/7/e34030
doi:10.2196/34030
PMID:35881418

Original Paper

High-Resolution Digital Phenotypes From Consumer Wearables and Their Applications in Machine Learning of Cardiometabolic Risk Markers: Cohort Study

Weizhuang Zhou[1*], PhD; Yu En Chan[1*], BSc; Chuan Sheng Foo[1], PhD; Jingxian Zhang[1], PhD; Jing Xian Teo[2], BSc; Sonia Davila[2,3,4], PhD; Weiting Huang[5], MBBS, MRCP; Jonathan Yap[5,6], MBBS, MRCP; Stuart Cook[4], MRCP, PhD; Patrick Tan[2,7,8,9], MD, PhD; Calvin Woon-Loong Chin[5,6], MD, MRCP; Khung Keong Yeo[2,5,6], MBBS; Weng Khong Lim[2,3,7], PhD; Pavitra Krishnaswamy[1], PhD

[1]Institute for Infocomm Research, Agency for Science Technology and Research (A*STAR), Singapore, Singapore
[2]SingHealth Duke-NUS Institute of Precision Medicine, Singapore, Singapore
[3]SingHealth Duke-NUS Genomic Medicine Centre, Singapore, Singapore
[4]Cardiovascular and Metabolic Disorders Program, Duke-NUS Medical School, Singapore, Singapore
[5]Department of Cardiology, National Heart Centre Singapore, Singapore, Singapore
[6]Duke-NUS Medical School, Singapore, Singapore
[7]Cancer and Stem Biology Program, Duke-NUS Medical School, Singapore, Singapore
[8]Cancer Science Institute of Singapore, National University of Singapore, Singapore, Singapore
[9]Genome Institute of Singapore, Agency for Science Technology and Research (A*STAR), Singapore, Singapore
[*]these authors contributed equally

Corresponding Author:
Pavitra Krishnaswamy, PhD
Institute for Infocomm Research
Agency for Science Technology and Research (A*STAR)
1 Fusionopolis Way, #21-01
Connexis (South Tower)
Singapore, 138632
Singapore
Phone: 65 64082450
Email: pavitrak@i2r.a-star.edu.sg

Abstract

Background: Consumer-grade wearable devices enable detailed recordings of heart rate and step counts in free-living conditions. Recent studies have shown that summary statistics from these wearable recordings have potential uses for longitudinal monitoring of health and disease states. However, the relationship between higher resolution physiological dynamics from wearables and known markers of health and disease remains largely uncharacterized.

Objective: We aimed to derive high-resolution digital phenotypes from observational wearable recordings and to examine their associations with modifiable and inherent markers of cardiometabolic disease risk.

Methods: We introduced a principled framework to extract interpretable high-resolution phenotypes from wearable data recorded in free-living conditions. The proposed framework standardizes the handling of data irregularities; encodes contextual information regarding the underlying physiological state at any given time; and generates a set of 66 minimally redundant features across active, sedentary, and sleep states. We applied our approach to a multimodal data set, from the SingHEART study (NCT02791152), which comprises heart rate and step count time series from wearables, clinical screening profiles, and whole genome sequences from 692 healthy volunteers. We used machine learning to model nonlinear relationships between the high-resolution phenotypes on the one hand and clinical or genomic risk markers for blood pressure, lipid, weight and sugar abnormalities on the other. For each risk type, we performed model comparisons based on Brier scores to assess the predictive value of high-resolution features over and beyond typical baselines. We also qualitatively characterized the wearable phenotypes for participants who had actualized clinical events.

Results: We found that the high-resolution features have higher predictive value than typical baselines for clinical markers of cardiometabolic disease risk: the best models based on high-resolution features had 17.9% and 7.36% improvement in Brier score over baselines based on age and gender and resting heart rate, respectively (P<.001 in each case). Furthermore, heart rate dynamics from different activity states contain distinct information (maximum absolute correlation coefficient of 0.15). Heart rate dynamics in sedentary states are most predictive of lipid abnormalities and obesity, whereas patterns in active states are most predictive of blood pressure abnormalities (P<.001). Moreover, in comparison with standard measures, higher resolution patterns in wearable heart rate recordings are better able to represent subtle physiological dynamics related to genomic risk for cardiometabolic disease (improvement of 11.9%-22.0% in Brier scores; P<.001). Finally, illustrative case studies reveal connections between these high-resolution phenotypes and actualized clinical events, even for borderline profiles lacking apparent cardiometabolic risk markers.

Conclusions: High-resolution digital phenotypes recorded by consumer wearables in free-living states have the potential to enhance the prediction of cardiometabolic disease risk and could enable more proactive and personalized health management.

(J Med Internet Res 2022;24(7):e34669) doi:10.2196/34669

KEYWORDS

wearable device; heart rate; cardiometabolic disease; risk prediction; digital phenotypes; polygenic risk scores; time series analysis; machine learning; free-living

Introduction

Background

The adoption of consumer-grade wearable activity trackers into routine use has been increasing rapidly in recent years, with approximately 1 in 5 adults in the United States reported to regularly use wrist-worn smartwatches and fitness trackers in 2019 [1]. This phenomenon has generated an unprecedented scale of consumer health data and led to many studies on the wider health uses of such data. These studies are increasingly generating evidence to reveal relationships between recordings from wearable activity trackers and the risk for conditions ranging from mental health and infectious diseases [2,3] to cardiovascular and metabolic (collectively referred to as *cardiometabolic*) diseases [4-7]. Among these, owing to the apparent links between activity levels and cardiometabolic health, the evidence for broader health uses of wearables is most established in the cardiometabolic domain [4,8-11].

Previous studies in the cardiometabolic domain have focused on the utility of wearable-derived summary statistics, and fall into 1 of 2 categories. First, electrocardiogram signals from wearables have been studied in relation to the development of cardiometabolic conditions, such as atrial fibrillation [12-14], hyperkalemia [15,16], and heart failure [17-19]. As many of these conditions are amenable to early intervention via dietary changes or increased physical activity, there is also an interest in using wearables to promote self-awareness and regulation [20] and to enhance screening [11]. Second, wearable-derived measures, such as circadian measures, sleep patterns and quality [11,21], step counts [4], wearable-derived resting heart rate [4,8,10,21,22] and heart rate variability [23-27] have been found to correlate with outcomes in cardiometabolic disease. As such, there is increasing recognition in the clinical community to incorporate wearable-derived measures into practical cardiometabolic disease management [6,28].

Objectives

Rapid and ongoing developments in consumer wearable technology are enabling ever-richer measurements with finer temporal resolution for heart rate, activity, and sleep dynamics in free-living states [6,29,30]. Principled analyses of such data streams could generate new insights beyond summary statistical measures for cardiometabolic health and disease management. However, the analysis of time series data recorded in free-living states is challenging, as these data tend to exhibit real-world noise and fluctuations and typically lack important physical and physiological contexts. A few recent studies have used black-box deep neural networks to relate high-resolution heart rate and step count time series recorded using wearables to the risk of developing atrial fibrillation, sleep apnea, and hypertension [31,32]. As their primary goal focused on risk target classification, the nature of the intermediate predictive time series features and their connection with known clinical and biological markers of cardiometabolic disease remains unresolved.

In this study, we aimed to derive high-resolution digital phenotypes from consumer wearable heart rate recordings and to examine their associations with diverse risk markers for cardiometabolic disease. Specifically, we sought to develop a time series feature extraction approach, contextualized by activity state, to meaningfully represent heart rate dynamics recorded by consumer wearables in free-living conditions. We then applied our approach to multidimensional data from normal volunteers in the SingHEART study [33] to assess the extent to which the derived high-resolution wearable features could predict expressed clinical risk markers for cardiometabolic disease. Furthermore, we assessed whether these high-resolution features also represent more subtle physiological changes associated with an inherent genetic predisposition to cardiometabolic disease. Finally, we qualitatively characterized these wearable phenotypes in volunteers who had actualized clinical events to assess connections beyond risk markers to manifest cardiometabolic diseases.

Methods

Data

We sourced data from the SingHEART study (NCT02791152) as of October 8, 2019. Enrollment targeted healthy volunteers who provided written informed consent to use the data (including electronic health records) for research. Participants were required to fulfill the inclusion criteria presented in Textbox 1.

Textbox 1. Inclusion criteria.

Inclusion criteria

1. 21-69 years of age
2. No personal medical history of prior cardiovascular disease (myocardial infarction, coronary artery disease, peripheral arterial disease, stroke), cancer, autoimmune or genetic disease, endocrine disease, diabetes mellitus, psychiatric illness, asthma, chronic lung disease, or chronic infectious disease
3. No family medical history of cardiomyopathies

At enrollment, each participant was profiled using a range of health assessment modalities. The resulting data set included (1) heart rate and step count time series recordings over 3 to 5 days from consumer wearable devices (Fitbit Charge HR), together with the associated sleep logs generated by Fitbit, (2) self-reported answers to a lifestyle and quality-of-life questionnaire [4], (3) genotypic data from whole genome sequencing using the Illumina HiSeq X platform, and (4) laboratory measurements for 9 clinically relevant markers (systolic and diastolic blood pressure; blood levels of triglycerides, total cholesterol, high-density lipoprotein, and low-density lipoprotein; fasting blood glucose level; waist circumference and BMI). As of October 8, 2019, the full study cohort contained 1101 participants, of whom 692 (62.8%) participants had wearable recordings. We focused on this subset of participants for subsequent analysis: a detailed breakdown of the data is provided in Table 1.

Furthermore, we also tracked each participant for the occurrence of any actual clinical event. We extracted all clinical codes (based on the International Classification of Diseases, 10th Revision) pertaining to any acute care use events in the regional health system associated with the National Heart Centre Singapore until January 2021 to characterize the links among data features, risk markers, and actual clinical events.

Table 1. Summary of demographic, clinical, and consumer wearable data for participants with wearable recordings (N=692) in the SingHEART study cohort.

	Female (n=370, 53.5%)		Male (n=322, 46.5%)	
	Value, mean (SD)	Participants, n[a] (%)	Value, mean (SD)	Participants, n[a] (%)
Age (years)	45.47 (11.71)	0 (0)	44.46 (13.29)	0 (0)
BMI (kg/m^2)	22.87 (3.94)	0 (0)	24.33 (3.39)	0 (0)
WC[b] (cm)	78.91 (10.98)	0 (0)	86.96 (9.86)	0 (0)
SBP[c] (mm Hg)	122.51 (17.74)	0 (0)	132.20 (14.96)	0 (0)
DBP[d] (mm Hg)	73.38 (12.80)	0 (0)	82.18 (10.97)	1 (0.3)
Wearable-derived resting heart rate (bpm; Fitbit)	70.66 (6.55)	0 (0)	69.37 (6.59)	0 (0)
ECG_HR[e] (bpm)	64.46 (9.17)	10 (2.7)	63.67 (9.87)	12 (3.7)
Total cholesterol (mmol/L)	5.34 (0.94)	6 (1.6)	5.33 (0.97)	5 (1.6)
LDL[f] (mmol/L)	3.32 (0.81)	7 (1.9)	3.40 (0.89)	6 (1.9)
HDL[g] (mmol/L)	1.59 (0.32)	6 (1.6)	1.36 (0.30)	5 (1.6)
TGs[h] (mmol/L)	0.99 (0.51)	6 (1.6)	1.30 (0.76)	5 (1.6)
Glucose (mmol/L)	5.17 (0.49)	8 (2.2)	5.36 (0.71)	5 (1.6)
Average daily step count[i]	10,349.81 (4180.35)	30 (8.1)	10,972.86 (3919.10)	20 (6.2)
Average daily sedentary minutes	633.45 (96.48)	102 (27.6)	656.49 (95.58)	88 (27.3)
Average daily sleep minutes	395.92 (61.18)	102 (27.6)	374.49 (65.15)	88 (27.3)

[a]Refers to number of participants with missing or incomplete values for the respective fields.

[b]WC: waist circumference.

[c]SBP: systolic blood pressure.

[d]DBP: diastolic blood pressure.

[e]ECG_HR: electrocardiogram heart rate.

[f]LDL: low-density lipoprotein.

[g]HDL: high-density lipoprotein.

[h]TG: triglyceride.

[i]The average daily step count was derived by taking the sum of steps for each day and then averaging over days. Only days with ≥20 hours of valid data were considered.

Ethics Approval

The SingHEART study (NCT02791152) was established at the National Heart Centre Singapore, a tertiary specialty hospital in Singapore, and was approved by the SingHealth Centralized Institutional Review Board (ref: 2015/2601 and 2018/3081) [33,34].

A Set of 22 Canonical Time Series Characteristics

Given a time series segment, it is possible to define a set of high-resolution features using approaches such as the highly comparative time series analysis [35,36] and time series feature extraction on the basis of scalable hypothesis [37,38]. However, such approaches can generate many redundant features, and the process of selecting a concise but effective representation is often not straightforward. A recent study [39] introduced a minimally redundant and interpretable set of 22 features, termed as Canonical Time-series Characteristics 22 (Catch22) features, which have high predictive value across 93 diverse time series classification data sets. As this Catch22 feature set was designed to reduce interfeature redundancy, it provides a compendious representation of the different dynamic properties of the time series.

The Catch22 features fall into seven main categories, namely (1) distribution, (2) extreme events, (3) symbolic, (4) linear autocorrelation and periodicity, (5) nonlinear autocorrelation, (6) successive differences, and (7) fluctuation analysis. The distribution-based features represent summary statistics of the distribution of the measured values in the series (while ignoring the chronological order of these values). The extreme event features represent intervals between successive outlier events in the time series. The symbolic features represent statistics summarizing the outputs of symbolic transformations of the actual time series values. The linear autocorrelation and periodicity features comprise summary statistics on inherent periodicities in the time series. The nonlinear autocorrelation features involve summary statistics on periodicities based on nonlinear transformations of the time series. The successive difference features represent statistics based on the time series of the incremental differences. Finally, the fluctuation analysis

features quantify the statistical self-affinity of the time series. Detailed descriptions of each of the 22 features are provided in Table S1 in Multimedia Appendix 1.

Extraction of Features From Wearable Time Series Recordings

We now describe the steps to derive resting heart rate, summary statistics on activity and sleep patterns, and high-resolution features from the wearable heart rate and step count time series recordings. As all these physiological features are derived from the same recordings, they are internally consistent and can be meaningfully used for downstream comparative analyses.

Computation of RestingHR

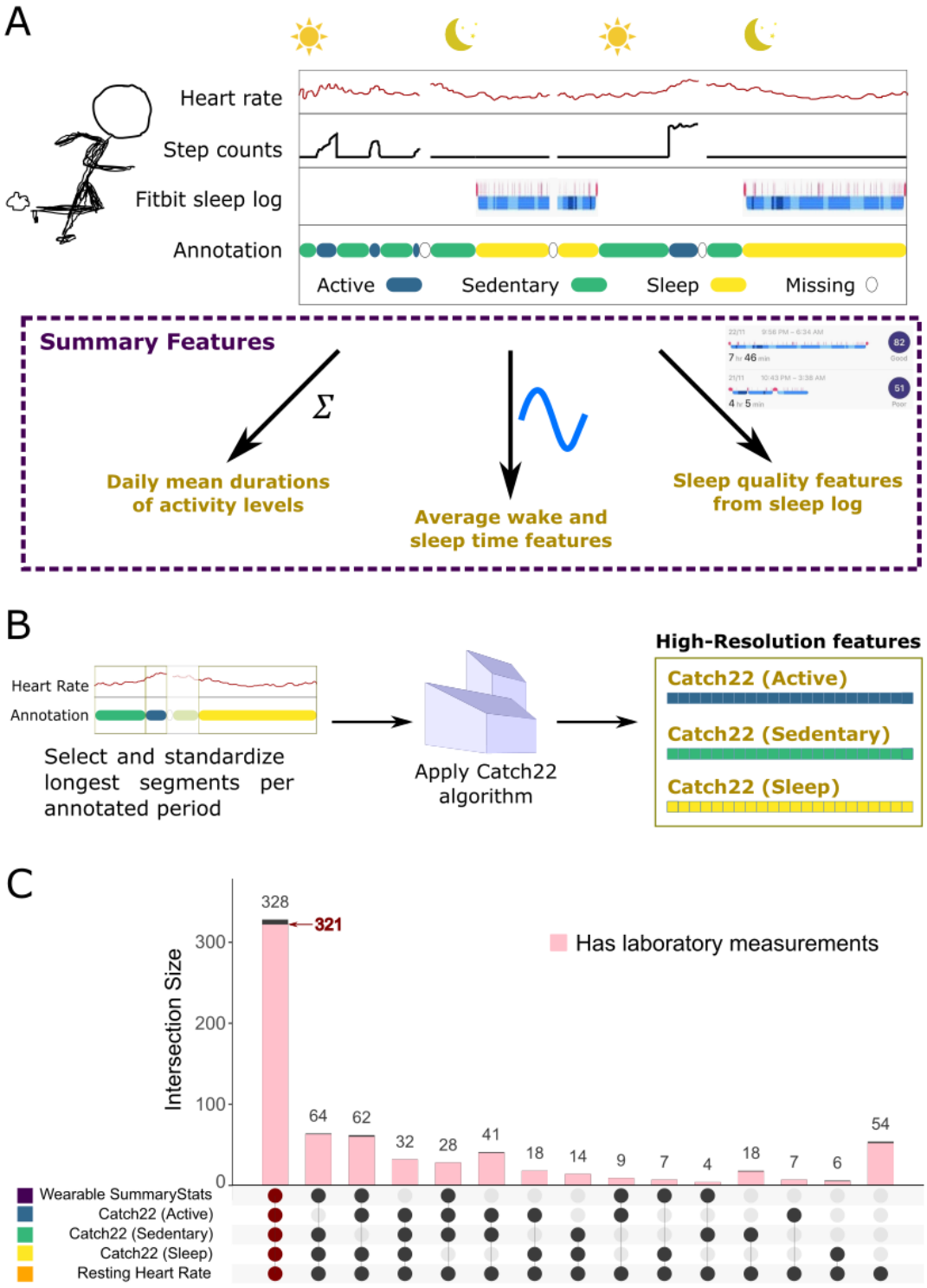
We used wearable heart rate time series recordings to derive resting heart rate [4]. Specifically, we defined *wearable-derived resting heart rate* as the average of heart rate values across all time points that had a valid heart rate record and a step count of ≤100. We note that there are similarities between the wearable-derived resting heart rate and the clinical gold standard electrocardiogram-derived heart rate [4,40].

Annotation of Wearable Time Series Recordings

We extracted the wearable time series recordings for each participant and used only days with at least 20 hours of step count and heart rate data as per Lim et al [4]. Heart rate recordings were available either at regular 1-minute intervals or as irregular bursts of recordings over 5-, 10-, or 15-second intervals. Step count recordings were sampled at either 15-minute or 1-minute intervals. We resampled all heart rate and step count consumer wearable records to 1-minute intervals and then annotated the time series to reflect data availability and physical activity states (Figure 1A). We assigned a *null* value for heart rate at time points where it was missing. Then, we annotated time points with available data for both heart rate and step count as "sleep," "active," or "sedentary." Specifically, we applied the *sleep* annotation to all time points captured by the Fitbit sleep log, the *sedentary* annotation to any time points with 0 step count value, and denoted the remaining time points as *active*. On average, the participants in our study had 3.72 days of valid heart rate data, and the average missing heart rate periods in a day were 94.9 (SD 85.8) minutes. The median lengths of the longest uninterrupted time series for the *active*, *sedentary*, and *sleep* periods were 31, 105, and 465 minutes, respectively.

Subsequently, we processed the heart rate and step count time series recordings from the consumer wearable devices to yield a range of summary and high-resolution features, as detailed in subsections *Derivation of Summary Features From Wearable Time Series Recordings* and *Derivation of High-Resolution Features From Wearable Time Series Recordings*.

Figure 1. Wearable data processing pipeline. (A) Construction of low-resolution features based on summary statistics. (B) Construction of high-resolution features based on the Canonical Time-series Characteristics 22 (Catch22) algorithm. (C) UpSet plot of the 692 participants with features from the various categories. Only nonempty set intersections are presented. Intersection size indicates the number of participants found within the intersections of given sets. Of the largest intersection with 328 participants, 321 also had laboratory measurement recordings.



Derivation of Summary Features From Wearable Time Series Recordings

We used a 3-step procedure to derive a range of wearable summary statistics (Figure 1A). First, we used our physical activity annotations to compute mean daily durations for the different activity states. Second, we used device logs to obtain statistics relating to sleep-wake patterns. Third, we converted the wake and sleep times into a 24-hour format and averaged the resulting values over all days where a given participant had wearable data recordings. To account for the cyclical nature of sleep or wake patterns, we transformed the average wake and sleep times using sinusoidal functions. Overall, this process

yielded 10 summary features for each participant. All summary statistics are listed in Table S2 in Multimedia Appendix 1.

Derivation of High-Resolution Features From Wearable Time Series Recordings

We further developed a data processing pipeline to extract high-resolution time series features from heart rate recordings of the wearable device (Figure 1B). As heart rate and step count patterns under different physical activity states could provide distinct insights into cardiovascular health, we sought to derive time series features that encode contextual information about the physical activity state. Specifically, we processed heart rate time series recordings for each of the 3 physical activity states (sleep, sedentary, and active) separately, as follows.

For each participant, we chose the longest uninterrupted period of the heart rate time series recordings for each physical activity state. As the data exhibit significant variability in the lengths of these periods across participants, we defined prespecified lengths to extract standardized sleep, sedentary, and active segments. Specifically, we extracted the first 20 minutes for active segments, the first 1 hour for sedentary segments, and the first 5 hours for sleep segments. If the recordings available for a participant did not fulfill the prespecified length criteria, even with the longest segment for a given activity state, we did not consider that particular activity state for high-resolution analyses. This process yielded up to 3 heart rate time series segments for each participant.

For each available heart rate time series segment, we applied the Catch22 methodology [39] to obtain 22 high-resolution features. Collectively, our pipeline resulted in up to 3 sets of 22 high-resolution features per participant, namely Catch22 (Sleep), Catch22 (Active), and Catch22 (Sedentary).

As our study did not prescribe controlled experimental settings for the wearable recordings, the resulting time series segments often exhibit significant noise and irregularities. Hence, we considered the reliability of our feature representation approach in these real-world settings. In particular, we assessed stability and sensitivity of the Catch22 features to the length specifications across activity states (Section SI-1, Multimedia Appendix 1). The results suggest that the features are relatively robust within the intervals considered and provide confidence for the downstream use of these high-resolution features.

Overlap Among Features Derived From Wearable Time Series Recordings

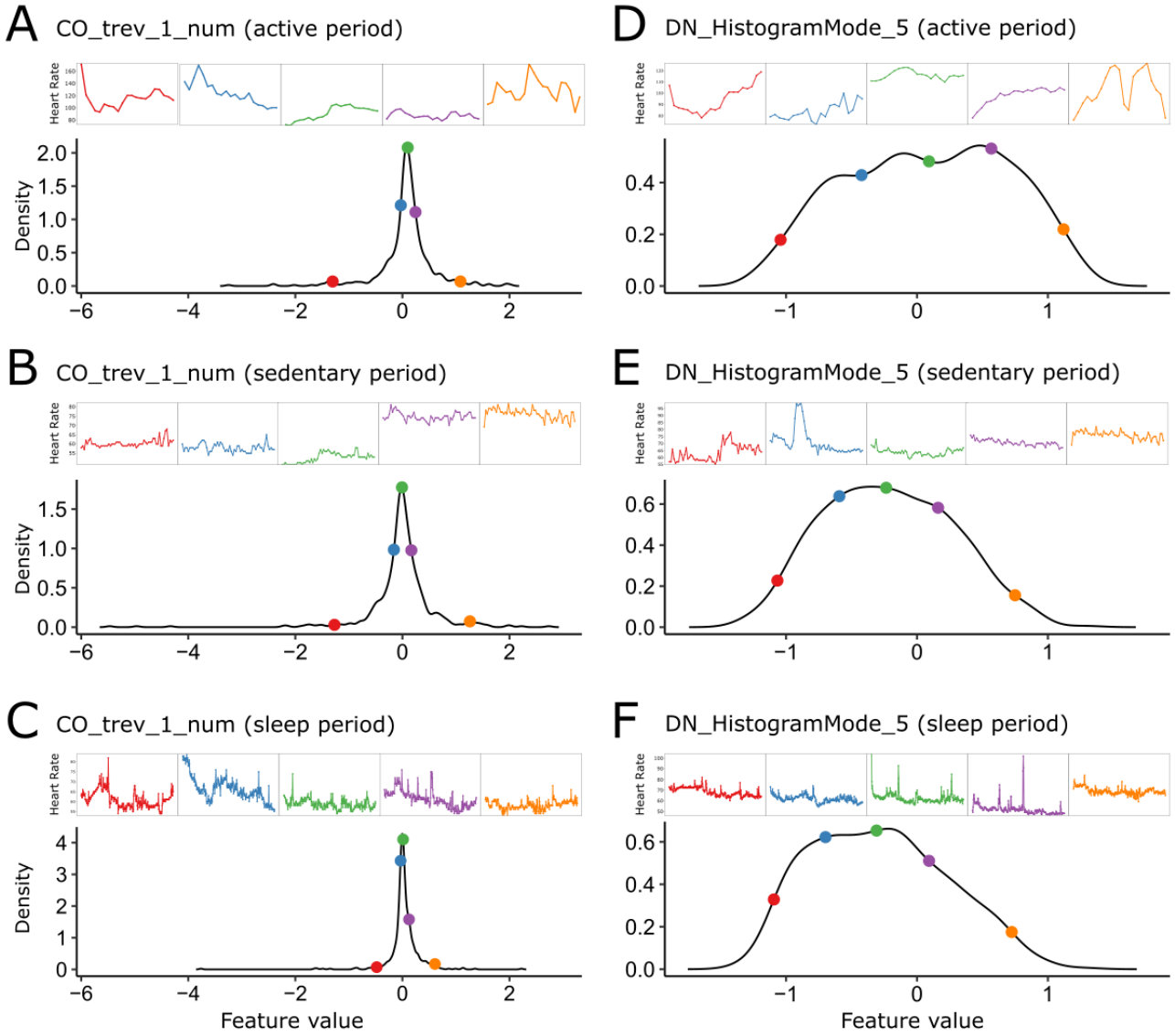
Figure 1C illustrates the overlaps among participants with the different wearable-derived features using UpSet plots [41,42]. For example, 41 individuals had features for active and sedentary segments but did not have sleep segments or summary statistics (owing to a lack of sufficiently long continuous sleep recordings). We note that all the different types of wearable features are available for 328 participants, of which 321 (97.9%) also had laboratory measurements. We considered this set of 321 participants for ensuing visualization, risk modeling, and analysis.

Visualization of High-Resolution Heart Rate Features From Wearables

We examined how high-resolution wearable-derived heart rate features from sleep, active, and sedentary segments were distributed across study participants. Figure 2 illustrates the empirical distributions of exemplar features drawn from segments corresponding to each of the 3 physical activity states. To examine the variability across participants, we also visualized representative time series at the 2.5th, 25th, 50th, 75th, and 97.5th percentile of the density.

The first example comprises a nonlinear autocorrelation feature (CO_trev1_num, quantifying the time-reversibility statistic $<(x_{t+1}-x_t)^3>_t$) that relates to the degree of spikiness or regularity in the wearable-based heart rate time series (Figures 2A-2C). The second example comprises a distribution feature (DN_HistogramMode_5, corresponding to the mode of the z-transformed values) that quantifies the degree of nonnormality of the time series values by representing the difference between the most probable values (mode) and the mean of the series (Figures 2D-2F).

Figure 2. Illustration of wearable-derived high-resolution heart rate features. The distributions of 6 high-resolution features from the 321 participants, based on 2 Canonical Time-series Characteristics 22 features obtained from time series recordings in each of the 3 activity levels. The selected participants are at the 2.5th, 25th, 50th, 75th and 97.5th percentiles of each distribution, and the time series for the participant is plotted in the corresponding color. (A-C) CO_trev1_num is the time-reversibility statistic; higher values tend to correspond to "spikier" or irregular time series. (D-F) DN_HistogramMode_5 takes a time series and groups the z-scored values into 5 linearly spaced bins and reports the mode of the bins.



Characterization of Predictive Value of Wearable-Derived Features for Clinical Targets

Overview

The overall approach used to characterize the predictive value of different wearable-derived features with respect to a variety of clinical risk markers is as follows. Specifically, we considered model types based on 6 different feature sets (Table 2). We then defined 4 target clinical risk markers based on whether the 9 laboratory measurements exceeded the thresholds in Table 3: (1) abnormal blood pressure readings (*bp_abnormal*); (2) abnormal lipid levels (*lipids_abnormal*) for at least one of III to VI; (3) obese (*obesity*) for either VIII or IX; and (4) an omnibus category for lipid, blood sugar, obesity and sugar abnormalities (*anyRISKoutof9*) for any of I to IX.

All 321 participants who had a complete set of wearable-derived features also had complete data for the 9 laboratory measurements. We considered this set of 321 participants as our training set to model the clinical risk targets. Of these 321 participants, 149 (46.4%) were not positive for any of the 4 risk markers, whereas 172 (53.5%) were positive for at least one risk marker (Section SI-2, Multimedia Appendix 1). We noted that a given participant can be positive for >1 of the 4 labels, but most participants exhibiting positive risk markers were exclusively labeled by a single risk marker. Of the 172 positive participants, 119 (69.2%) were positive for 1 clinical risk marker, 40 (23.3%) were positive for 2 risk markers, and only 14 (8.1%) were positive for 3 or more risk markers.

Table 2. Description of the different model types.

Model name	Features included	Features, n
Baseline [4]	Age+gender	2
RestingHR	Baseline features+wearable-derived resting heart rate	3
SummaryStats	Baseline features+wearable summary stats	12
HighRes.ActiveSeg	Baseline features+Catch22[a] (active)	24
HighRes.SedenSeg	Baseline features+Catch22 (sedentary)	24
HighRes.SleepSeg	Baseline features+Catch22 (sleep)	24

[a]Catch22: Canonical Time-series Characteristics 22.

Table 3. Laboratory measurements and corresponding thresholds.

Laboratory measurement	Threshold to be considered at risk
I. Systolic blood pressure (mm Hg)	>140
II. Diastolic blood pressure (mm Hg)	>90
III. Triglycerides (mmol/L)	>2.3
IV. Total cholesterol (mmol/L)	>6.2
V. HDL[a] (mmol/L)	<1
VI. LDL[b] (mmol/L)	>4.1
VII. Fasting blood glucose level (mmol/L)	>6
VIII. Waist circumference (cm)	
Male	>100
Female	>90
IX. BMI (kg/m^2)	>27.5

[a]HDL: high-density lipoprotein.

[b]LDL: low-density lipoprotein.

We used machine learning to model the complex nonlinear relationships between a given feature set and the target pairing using 2 separate approaches. First, for any given target, we analyzed the predictive value of different feature sets (Table 1) using a model comparison approach. Specifically, we consider the degree to which the wearable-derived features (resting heart rate, wearable summary statistics, and different high-resolution wearable features) augment the predictive value of the baseline demographic feature set and also compared the performance of the high-resolution wearable features with that of the lower-resolution features. For an appropriate comparison of value addition over the baseline features, all feature sets based on wearable data also included the corresponding baseline features. Second, for each prediction target, we also compared the importance of the individual feature variables. To have a common basis for these variable importance calculations, we developed a unified model with all features included, and used this model to compare variable importance for the different features.

Prediction Model and Variable Importance

We trained machine learning models to estimate the probability that a participant exhibits clinical risk markers for common cardiometabolic disease abnormalities. Specifically, we used random forest classifiers [43] to model the 4 targets of interest, as they are general purpose, nonlinear classifiers that perform well in diverse settings. We trained the random forest models in R using the *randomForest* package [44]. To handle the imbalanced nature of the prediction tasks at hand, we set the number of minority class samples chosen for each tree at 80% of the total minority class size. We then down-sampled the majority class to match the number of minority class samples used [45]. This was implemented using the *strata* and *sampsize* parameters. For each of the 4 prediction targets, we constructed 200 such random forests with different starting random seeds, and for each random forest trained, we recorded the out-of-bag (OOB) prediction errors.

For random forests, variable importance can be quantified using the mean decrease in accuracy (MDA) over all OOB cross-validated predictions. To obtain statistically robust estimates of variable importance, for a given prediction target, we averaged the MDA for each feature across the 200 random forests and then ranked the features by their average MDA to obtain the top 10 important features. To visualize the variable importance results, we considered the union of the top 10 ranking features for the 4 cardiometabolic disease risk targets.

Model Performance Metric and Assessment

As the risk prediction task is inherently probabilistic, a suitable metric for model performance assessment would emphasize the calibration of the model predictions (ie, the prediction probabilities of true positives and true negatives are close to 1 and 0, respectively). Therefore, we evaluated the accuracy of probabilistic predictions using the Brier score [46]:

$$\text{BrierScore}(M) = [\sum_{i=1}^{N} (p_i - o_i)^2] / N \quad \textbf{(1)}$$

where M is the wearable-based model under consideration, p_i is the prediction probability of observing target i using the model under evaluation, o_i is the actual observed target or label (binary:0/1), and N is the total number of participants included in the modeling. The Brier score ranges from 0 to 1 and is lower for models with better calibrated predictions.

We used OOB estimates [43,47] to evaluate the scores, as there were insufficient data for an independent held-out test set. In total, the above process yielded 200 Brier scores for each pairing of the prediction target and wearable-derived feature set (model) type.

For each target, we also compared the performance of the various model types in relation to each other. Specifically, for each pair of model types, we performed a 2-tailed Welch t test on the null hypothesis that the true difference in Brier scores was 0. For each target, we corrected for multiple hypothesis testing by controlling the false discovery rate [48].

Characterization of Associations Between Wearable-Derived Features and Genomic Risk Markers

To better understand the nature of wearable-derived time series features, we investigated their associations with genomic risk markers for cardiometabolic disease. As probing these associations requires handling diverse multidimensional data types with potentially complex nonlinear relationships, we used a machine learning framework (similar to the one described earlier) to model these relationships. We then used model performance measures to infer the degree of information overlap between wearable features and genomic risk targets. As genomic risk is independent of age, we did not include age in any of the models considered.

We categorized the genetic susceptibility to cardiometabolic diseases using polygenic scores (PGSs). To define the genomic risk for lipid abnormalities, blood pressure abnormalities, and obesity, we used the PGS Catalog [49] to identify relevant polygenic risk scores corresponding to the 3 targets. Specifically, we identified 14 PGS for lipid abnormalities (PGS000060, PGS000061, PGS000062, PGS000063, PGS000065, PGS000115, PGS000192, PGS000309, PGS000310, PGS000311, PGS000340, PGS000677, PGS000688, and PGS000699), 2 for blood pressure abnormalities (PGS000301 and PGS000302), and 1 for obesity (PGS000298). Additional details of the selection process are provided in Section SI-3 (Multimedia Appendix 1).

For each of the 3 targets, we labeled a participant as having high genomic risk if their scores for any of the relevant PGS were in the top or bottom decile (refer to Section SI-3, Multimedia Appendix 1 for how the direction of a PGS is determined), which we term as the 90/10 cut-off. For instance, the high genomic risk group for lipid abnormalities would include members with high-risk scores for at least 1 of the 14 lipid-related PGS. The modeling of these targets and statistical comparison of the performance of different model types were identical to the earlier process described for the clinical risk targets.

To evaluate the sensitivity of the chosen percentile cut-offs for genomic risk scores, we repeated the above analyses for 2 additional sets of cut-offs, namely the 80/20 and 85/15 cut-offs.

Illustrative Profiling Based on Clinical Events

Finally, we examined the connections between high-resolution wearable-derived features and actualized cardiometabolic disease events for participants not in our training set of 321. Among these participants, we considered those who actualized cardiometabolic disease events indicated by a primary diagnosis of cardiovascular disease, dyslipidemia, and hypertension (as per International Classification of Diseases, 10th Revision codes listed in Table S3 in Multimedia Appendix 1). As this set of events spans a broad range of cardiometabolic conditions, anyRISKoutof9 is the closest surrogate marker. Hence, we chose to focus our profiling on the wearable-derived feature set that was most strongly associated with anyRISKoutof9.

For participants selected per the abovementioned criteria, we examined demographic information, physical measurements, genomic risk of disease, and clinical risk markers alongside the wearable-derived features. To interpret how the different wearable-derived features contribute to the model predictions at the individual participant level, we computed the Shapley values (Φ) [50] of each feature using the *iml* package [51] in R and selected the 5 features with the highest absolute magnitude of Φ for each participant. We illustrate the profiles of the participants, the predictions made by the best-performing model for anyRISKoutof9, and the features that contribute most to these predictions for each selected participant.

Software and Code Availability

All statistical analyses and modeling were performed using R Statistical Software (version 4.0.3; R Core Team 2020). Computation of resting heart rate was performed using R, but all other feature engineering efforts such as annotation of wearable time series recordings and derivation of summary features, as well as the generation of high-resolution features, were performed using Python (version 3.8.6).

All Python and R codes used in feature generation are available in Multimedia Appendix 2.

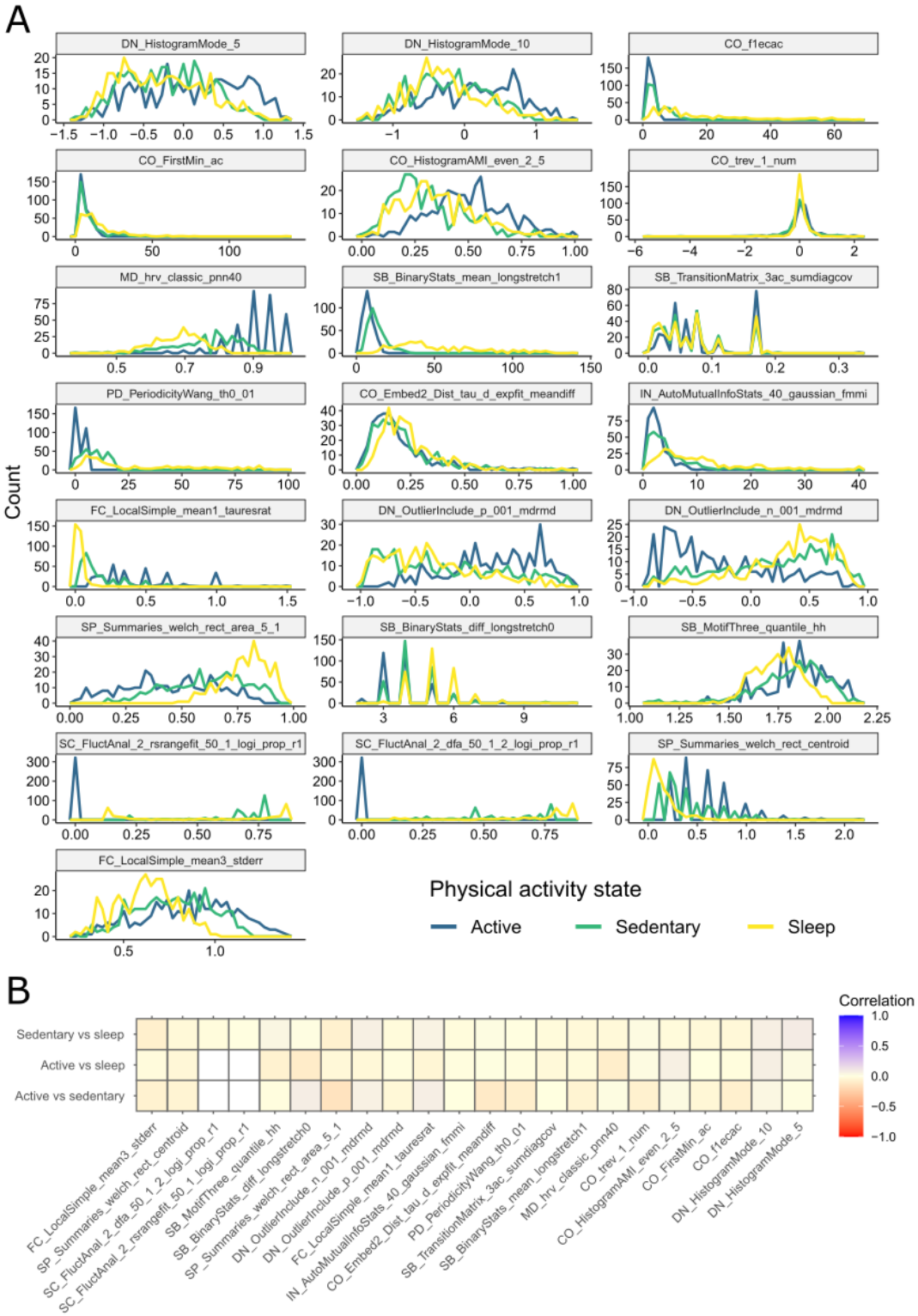
Results

Characteristics of High-Resolution Heart Rate Features From Wearables

Unlike summary statistics such as resting heart rate, which averages heart rate measurements across multiple days, our high-resolution feature sets provide more granularity on the

heart rate time series dynamics during different physical activity states (sleep, active, and sedentary). Figure 3A illustrates the distributions of the high-resolution wearable feature values across the 321 participants (colored according to their respective activity states). Although the Catch22 algorithm was identically applied to each of the 3 activity segments, we observed that each feature exhibited distinct distributions across the 3 different activity states.

Figure 3. High-resolution (Canonical Time-series Characteristics 22 [Catch22]) wearable features from 3 different activity states. (A) Frequency polygons of the feature values based on the training set. The colors indicate activity states. (B) Pearson correlation coefficients between pairs of Catch22 features from different physical activity states (sleep, active, and sedentary). Two features from the active period (SC_FluctAnal_2_rsrangefit_50_1_logi_prop_r1 and SC_FluctAnal_2_dfa_50_1_2_logi_prop_r1) are uniformly 0; hence, correlation coefficients involving these 2 features are undefined (white squares).



To study whether this difference holds at the participant level, we characterized the correlations among the high-resolution feature sets obtained during the 3 different activity states. For any given feature (eg, CO_trev1_num), we considered vectors of feature values for each physical activity state across the population (eg, CO_trev1_num.active, CO_trev1_num.sedentary, and CO_trev1_num.sleep). We then calculated the Pearson correlation between these feature vectors for each pair of the activity states. This analysis revealed that the feature values from the different activity states were poorly correlated (Figure 3B). In fact, the largest absolute correlation coefficient among any of the pairs was 0.15. Taken together, these findings indicate that heart rate dynamics from different activity states contain distinct information.

Predictive Value of Wearable-Derived Features for Clinical Targets

Having gained some intuition about the information contained within the wearable-derived feature sets, we considered their predictive value for the clinical markers of cardiometabolic disease risk. Specifically, we trained random forest models to use the different wearable-derived feature sets to classify each of the 4 cardiometabolic disease risk targets. We performed comparative analyses to evaluate the predictive value of the different wearable-derived feature sets for classification of the 4 cardiometabolic disease risk targets.

First, we compared the OOB performance of the models trained using different feature sets for each clinical risk marker target (Table 4). For each target, the best-performing model was based on one of the high-resolution wearable feature sets (HighRes.ActiveSeg, HighRes.SedenSeg, or HighRes.SleepSeg). Specifically, for anyRISKoutof9, the HighRes.SedenSeg model was the best-performing model, with 17.9% and 7.36% lower Brier scores than baselines based on age and gender and resting heart rate, respectively (P<.001 in each case). This finding highlights the predictive value of high-resolution information within wearable heart rate time series recordings.

Second, we observed that heart rate dynamics extracted from different activity level segments have differential predictive potential for the various targets, as evidenced by the statistically significant differences between Brier scores (P<.001) of the HighRes.ActiveSeg, HighRes.SedenSeg, and HighRes.SleepSeg models (Table 4). Of the 3 model types, HighRes.SedenSeg performs best for lipid abnormalities, obesity, and anyRISKoutof9, whereas HighRes.ActiveSeg performs best for blood pressure abnormalities.

Third, to comparatively evaluate contributions from individual wearable-derived features, we trained models that used all features available to predict each cardiometabolic disease risk target and ranked the variable importance in each case. Figure 4 shows the variable importance plots. It is clear that different features affect the performance of the models for each of the 4 targets. For instance, age and gender are the top 2 drivers of model performance for the anyRISKoutof9 target but are not among the top 10 features for both lipids_abnormal and obesity targets. Furthermore, we found that heart rate dynamics from different activity states contained distinct information on cardiometabolic disease risk. For example, the DN_HistogramMode_5 feature from the sedentary and active segments was important for predicting cardiometabolic disease risk markers but the DN_HistogramMode_5 feature from the sleep segment was not (Figure 4).

Fourth, we observed that the top 10 features for each of the 4 targets included features from all 6 feature types (age and gender, wearable-derived resting heart rate, wearable summary statistics, and the 3 sets of high-resolution features from Table 1). This suggests that risk prediction models using wearable-derived features may not exclusively rely on only one of the different feature sets or any one feature drawn from these feature sets. Rather, a collection of different wearable-derived high-resolution heart rate features from distinct activity states is essential for accurately predicting the multiplicity of cardiometabolic disease risk targets.

Table 4. Model performance on cardiometabolic risk targets. Out-of-bag model performance for each of the 5 model types computed for the 4 targets. A smaller Brier score indicates a better performing model for a given target.

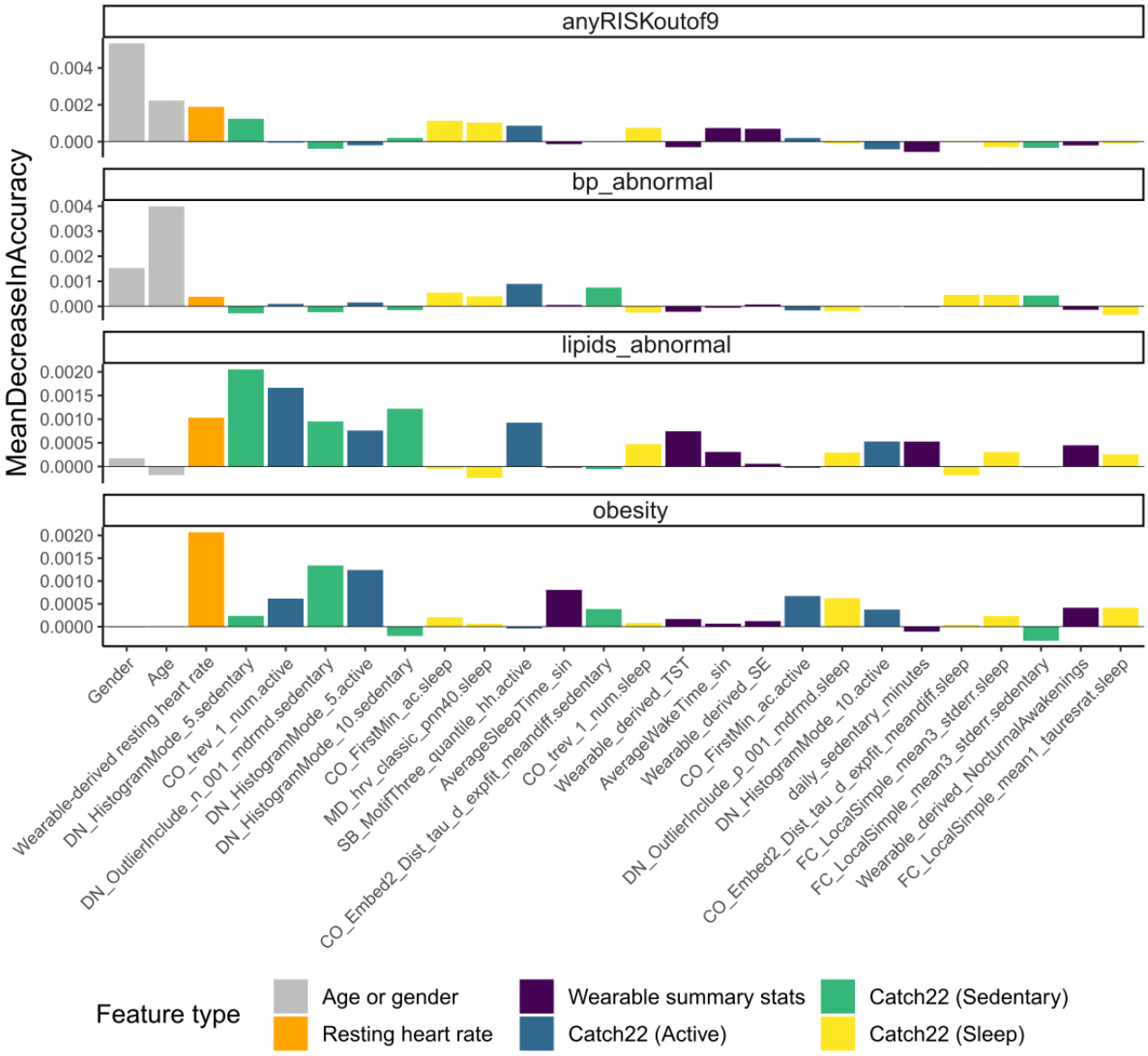
	Baseline[a], mean (SD)	RestingHR[b], mean (SD)	HighRes.ActiveSeg[c], mean (SD)	HighRes.SedenSeg[c], mean (SD)	HighRes.SleepSeg[c], mean (SD)	SummaryStats, mean (SD)
anyRISKoutof9	0.291 (-5.87×10^{-4})	0.258 (7.7×10^{-4})	0.253 (8.52×10^{-4})	0.239 (-9×10^{-4})	0.245 (8.43×10^{-4})	0.247 (7.66×10^{-4})
bp_abnormal	0.227 (4.79×10^{-4})	0.223 (5.61×10^{-4})	0.217 (7.88×10^{-4})	0.222 (8.14×10^{-4})	0.225 (8.32×10^{-4})	0.225 (7.9×10^{-4})
obesity	0.246 (6.64×10^{-4})	0.227 (7.91×10^{-4})	0.221 (8.92×10^{-4})	0.214 (9.34×10^{-4})	0.226 (8.64×10^{-4})	0.227 (8.54×10^{-4})
lipids_abnormal	0.271 (5.84×10^{-4})	0.261 (6.64×10^{-4})	0.238 (8.08×10^{-4})	0.225 (7.58×10^{-4})	0.241 (8.27×10^{-4})	0.236 (7.3×10^{-4})

[a]For each risk target, the Brier scores of the baseline model were significantly different from those of all other models (P<.001).

[b]For each risk target, Brier scores of the resting heart rate model (RestingHR) were significantly different from all other models (P<.001).

[c]For each risk target, Brier scores of the 3 HighRes models were significantly different from each other (P<.001).

Figure 4. Random forest variable importance. The variable importance of each feature for prediction of the 4 cardiometabolic disease risk targets. We averaged each importance value across 200 simulations and used the results to rank the top 10 features to retain for each cardiometabolic disease risk target. This resulted in a total of 26 features across all 4 targets, as shown in the figure. Catch22: Canonical Time-series Characteristics 22.



Associations Between Wearable-Derived Features and Genomic Risk Markers

To further interpret the information contained within the wearable-derived features, we sought to understand how they relate to the genetic predispositions for cardiometabolic diseases. Specifically, we examined the degree of information overlap between the different wearable-derived features (Table 1) and the genomic risk of cardiometabolic conditions. For each pairing between the different wearable-derived feature sets and the 3 genomic risk targets, we trained random forest models and used their Brier scores as indirect measures of the strength of the associations.

The results are presented in Table 5. For each of the 3 abnormality types, we observed that the high-resolution wearable features were more strongly associated with genomic risk levels than sex and resting heart rate (improvement of 11.9%-22.0% in Brier scores; P<.001). We highlight that the trends against baseline and resting heart rate were relatively insensitive to the polygenic risk score threshold used to define high versus low genomic risk (Section SI-4, Multimedia Appendix 1). These results suggest that, in comparison with standard measures, high-resolution features from wearables are better able to represent subtle physiological dynamics related to the genomic risk for cardiometabolic disease.

Table 5. Degree of association with genomic risk targets. Out-of-bag performance for each of the 5 model types computed for the 3 targets. A smaller Brier score indicates better performing model for a given target.

	Baseline[a], mean (SD)	RestingHR[b], mean (SD)	HighRes.ActiveSeg, mean (SD)	HighRes.SedenSeg, mean (SD)	HighRes.SleepSeg, mean (SD)	SummaryStats, mean (SD)
Blood pressure	0.248 (2.0×10^{-3})	0.245 (8.55×10^{-4})	0.215 (1.08×10^{-3})	0.214 (1.09×10^{-3})	0.215 (9.93×10^{-4})	0.212 (9.64×10^{-4})
Obesity	0.245 (2.31×10^{-3})	0.246 (9.03×10^{-4})	0.205 (1.15×10^{-3})	0.192 (1.06×10^{-3})	0.199 (1.21×10^{-3})	0.203 (1.06×10^{-3})
Lipids	0.294 (3.02×10^{-3})	0.308 (6.36×10^{-4})	0.254 (9.07×10^{-4})	0.254 (8.82×10^{-4})	0.259 (8.92×10^{-4})	0.268 (8.86×10^{-4})

[a]For each risk target, the Brier scores of the baseline model were significantly different from all other models (P<.001).

[b]For each risk target, Brier scores of the resting heart rate model (RestingHR) were significantly different from those of the 3 HighRes and SummaryStats models (P<.001).

Illustrative Profiles of Participants With Cardiometabolic Events

Finally, we examined the relationship between the wearable-derived feature set most predictive for anyRISKoutof9 and actualized cardiometabolic events. We focused on participants not in our training set and filtered participants with data for the feature set most predictive for anyRISKoutof9 (ie, Catch22 [Sedentary] feature set, based on the abovementioned results). This yielded 197 candidate participants for illustrative profiling. Among these participants, only 5 participants actualized events with primary diagnoses for cardiometabolic conditions (as specified in Table S3 in Multimedia Appendix 1).

Table 6 provides demographic, genetic, and clinical risk profiles along with physical measurements and important wearable features for these 5 participants (A-E). All the participants were aged 54 to 61 years. Of the 5 participants, 4 (80%) were male. Only 1 (20%) participant was obese. We now present the findings on the predictive value of high-resolution wearable-derived features for these participants.

First, we describe participants with abnormalities in both genetic and clinical risk markers, namely participants A and B. Participant A had high genomic risk for all 3 conditions, presented abnormal values for most of the 9 clinical risk markers, and was also diagnosed with all 3 types of cardiometabolic conditions considered (cardiovascular disease, dyslipidemia, and hypertension). Participant B had a genomic risk for lipid and blood pressure abnormalities, abnormal lipid panel values, and a clinical diagnosis of dyslipidemia. While participant A had a wearable-derived resting heart rate slightly above the population average, participant B had a wearable-derived resting heart rate lower than the population average. However, in both cases, our HighRes.SedenSeg model predicted a positive anyRISKoutof9 outcome.

Second, we considered participants with no genomic risk but who presented with abnormal clinical risk markers, namely participant C. This participant had high blood pressure, abnormal cholesterol and blood glucose levels, a clinical diagnosis of dyslipidemia, and wearable-derived resting heart rate slightly above the population average value. However, we noted that our HighRes.SedenSeg model predicted a negative anyRISKoutof9 outcome. This could be due to modeling error or possibly be attributed to the absence of severe changes in heart rate dynamics given the normal genetic background and moderate wearable-derived resting heart rate value.

Third, we highlighted participants who did not exhibit any abnormalities in clinical risk markers and were borderline for cardiometabolic disease risk, namely participants D and E. Participant D only had a genomic risk for blood pressure. Participant E, on the other hand, appeared to have the most benign profile with low genomic risk for all 3 target conditions and normal values for all 9 clinical risk markers (with only the BMI being borderline high). Both participants had wearable-derived resting heart rate values that were lower than the population average. Although participants D and E had a seemingly low-risk profile by standard measures, they had clinical diagnoses of dyslipidemia and cardiovascular disease, respectively. Indeed, our HighRes.SedenSeg model predicted a positive anyRISKoutof9 outcome in each case.

Finally, inspecting the most important features (top 5 Shapley values) contributing to model predictions for anyRISKoutof9 in Table 6 reveals interesting patterns. While age and gender were (expectedly) consistent contributors to prediction scores for most participants, many Catch22 (Sedentary) features also contributed at comparable levels. For instance, DN_Histogram_Mode_5 was important for all 5 participants, whereas CO_Embed2_Dist_tau_d_expfit_meandiff and DN_OutlierInclude_p_001_mdrmd were important for 3 and 2 participants, respectively. In particular, DN_Histogram_Mode_5 was an important feature for most participants in this study. This feature takes on large values when the participant's heart rate time series exhibits substantial deviations from the mean, which could occur when there are sustained or frequent oscillations with high amplitude. Although such deviations may be common in active states, their presence in sedentary states could forebode cardiovascular abnormalities, as was the case for these 5 participants. Beyond the consistent features noted above, there are other diverse high-resolution features among the top 5 most important contributors for different participants. This suggests that our high-resolution feature extraction approach offers a compact but sufficiently diverse set of predictive heart rate patterns, including those that are consistent across individual participants and those that can cater to participant-to-participant variability. Detailed Shapley Additive Explanations (SHAP) feature importance plots for each participant are provided in Section SI-5 in Multimedia Appendix 1.

Table 6. Illustrative profiles of 5 participants with actualized cardiometabolic events. Participant profiles include demographic information, type of cardiometabolic disease, key physical measurements, clinical and genomic risk markers, and the top 5 important wearable-derived heart rate features (as per Shapley values).

Participant profiles	Participant				
	A	B	C	D	E
Demographics					
Age (years)	54	57	56	55	61
Gender	Male	Male	Male	Female	Male
Wearable-derived resting heart rate	72.8	58.2	73.0	69.0	55.7
Clinical risk markers					
BMI (kg/m^2)	28.05	18.79	21.27	22.95	25.95
Blood pressure: SBP[a]/DBP[b] (mm Hg)	166/109	108/65	164/105	112/48	133/89
Glucose (mmol/L)	6.8	4.8	7.4	5.3	5.3
Total cholesterol (mmol/L)	5.27	6.63	6.60	5.05	4.45
anyRISKoutof9	True[c]	True	True	False[d]	False
High genomic risk					
Lipids abnormalities	True	True	False	False	False
Blood pressure abnormalities	True	True	False	True	False
Obesity	True	False	False	False	False
Actualized cardiometabolic events					
Cardiovascular disease	True	True	False	False	True
Dyslipidemia	True	False	True	True	False
Hypertension	True	False	False	False	False
Important features for prediction					
CO_f1ecac.sedentary	False	False	False	True	False
FC_LocalSimple_mean3_stderr.sedentary	True	False	False	False	False
SB_MotifThree_quantile_hh.sedentary	True	False	False	False	False
SB_TransitionMatrix_3ac_sumdiagcov.sedentary	False	False	False	True	False
CO_trev_1_num.sedentary	False	False	False	False	True
CO_HistogramAMI_even_2_5.sedentary	False	False	True	False	False
DN_OutlierInclude_p_001_mdrmd.sedentary	True	True	False	False	False
CO_Embed2_Dist_tau_d_expfit_meandiff.sedentary	False	True	False	True	True
DN_HistogramMode_10.sedentary	False	False	True	False	False
DN_HistogramMode_5.sedentary	True	True	True	True	True
Gender	True	True	True	False	True
Age (years)	False	True	True	True	True

[a]SBP: systolic blood pressure.

[b]DBP: diastolic blood pressure.

[c]True indicates true or that there is a presence of categorical variables.

[d]False indicates false or absence of categorical variables.

Discussion

Principal Findings

Consumer wearables enable the recording of rich high-resolution physiological dynamics in free-living conditions, but how these data relate to health and disease is not fully understood. We introduced a principled framework to derive high-resolution heart rate features from consumer wearable recordings, and applied our approach to a data set containing multidimensional cardiometabolic health parameters from healthy volunteers. Our results show that, in comparison with typical summary statistics,

high-resolution features resolving temporal dynamics and activity-dependent patterns in heart rate have stronger associations with modifiable risk markers and inherent genetic predispositions for cardiometabolic disease alike. Our findings imply that these high-resolution digital phenotypes from consumer wearables can provide a more granular picture of cardiometabolic health and disease states, which could have potential use in cardiometabolic health screening and disease management.

Our framework addresses key challenges in mining wearable data recorded in free-living conditions. Unlike clean data from controlled experimental settings, real-world wearable recordings tend to be irregular, contain missing stretches [29], lack clean context annotations, and have variable lengths. As such, analyses based on the naive application of general purpose time series feature extraction methods [36,39,52] may not have ecological validity [53]. To address this gap and derive meaningful physiological dynamics from wearable time series recordings, our feature extraction framework standardizes handling of data irregularities and encodes contextual information about the underlying activity level and physiological state (Figures 1-3). This conceptual framework, although demonstrated here with the Catch22 method [39], is agnostic to the choice of the feature representation method for time series data [36,37]. Furthermore, in contrast to black-box feature learning methods based on large labeled data sets [31], our approach yields more interpretable time series features with smaller unlabeled data sets.

Our framework provides many possibilities for gaining new insights from wearable recordings. Our analyses, using multimodal wearable, genomic, and clinical data from healthy volunteers, highlight 2 possibilities.

First, our results revealed new relationships between high-resolution heart rate dynamics from wearables and the risk of cardiometabolic disease. Most previous studies correlated clinically obtained measures of heart rate dynamics, such as heart rate variability, exercise capacity, and heart rate recovery, with disease risk or outcomes [54-56]. In contrast, our results revealed that heart rate dynamics recorded by consumer wearables, when processed appropriately, are also predictive of cardiometabolic disease risk (Tables 4-6; Figure 4). Furthermore, we found that heart rate dynamics from different activity states contain distinct information about specific cardiometabolic conditions (Table 4; Figures 3 and 4). For example, heart rate patterns in sedentary states are more related to abnormalities in lipid levels and obesity, whereas those in active states may be more related to abnormalities in blood pressure readings (Table 4). These findings highlight the value addition of assessing physiology in free-living activity states (beyond controlled clinical settings) for disease risk monitoring and management [57].

Second, our study provides new perspectives on the interrelations between wearable recordings and genetic predispositions in cardiometabolic diseases. Although there has been a longstanding interest in probing gene-lifestyle interactions and their additive effects on cardiovascular disease [58-60], such studies have had limited visibility on physiology in free-living conditions. We found surprising connections (Table 5) between high-resolution wearable-derived feature sets and genetic predispositions for cardiometabolic disease. As these associations did not appear to depend on the presence or absence of manifest clinical risk markers, we posit that high-resolution phenotypes from wearables may capture subtle subclinical physiological changes stemming from latent predispositions to disease.

Limitations

Although the uniquely multimodal nature of our data enables us to uncover many novel insights on high-resolution wearable phenotypes, limitations of data set size and cohort design present some challenges. First, it was infeasible to conduct full-scale gene-environment interaction studies [61-63]; or train state-of-the-art machine learning models with large feature sets. Second, as the risk of cardiometabolic disease is highly multifactorial, the limited visibility on relevant physical and lifestyle factors constrains the absolute predictive accuracy of all models presented. For instance, we had limited input on regular exercise habits as the observation span was less than a week, as well as limited overlap between key lifestyle indicators and wearable recordings (eg, only 9 participants who smoked had valid wearable records). Finally, as our study included only a small number of participants with actualized cardiometabolic events, we could not perform quantitative analyses to relate wearable phenotypes with clinical events. Future work based on larger cohorts [64] with more targeted study designs could address some of these limitations and enable cross-cohort validation of our current findings.

Conclusions

In conclusion, we demonstrated that high-resolution digital phenotypes based on heart rate patterns in wearable recordings provide important insights into physiology in free-living conditions. Our results revealed that these measures are associated with both genetic and clinical risk markers of cardiometabolic disease and have additional predictive value beyond wearable-derived summary statistics and clinical measures of cardiometabolic health. Hence, our work expands possibilities to use digital phenotypes from consumer wearables as readily accessible indicators of cardiometabolic health and disease and motivates new approaches for quantitative scoring of cardiometabolic disease risk. Future studies could expand our findings to even higher resolution digital phenotypes that can be extracted from recordings with newer generations of wearable devices [65,66] and target evaluations for precision screening, health monitoring, and disease management applications.

Acknowledgments

This research was supported by funding and infrastructure from the Singapore National Precision Medicine Program (IAF-PP H17/01/a0/007) and the Institute for Infocomm Research, A*STAR. Data acquisition was supported in part by funding from SingHealth, Duke-NUS Medical School, National Heart Centre Singapore, Singapore National Medical Research Council (NMRC/STaR/0011/2012, NMRC/STaR/0026/2015), Lee Foundation, and the Tanoto Foundation.

The authors would like to thank all the volunteers for their participation in this study. They acknowledge valuable data collection assistance from the National Heart Centre Singapore and SingHEART Clinical Research Coordinators and resources from the National Supercomputing Center, Singapore [67] for code development on a polygenic score computation pipeline. They also thank Marie Loh from Nanyang Technological University, Singapore, and Xueling Sim from the National University of Singapore for meaningful discussions on polygenic risk scores.

JZ was affiliated with the Institute of Infocomm Research at the time of his contribution to this work, and is currently affiliated with the Diagnostics Development Hub (DxD Hub) at the Agency for Science Technology and Research (A*STAR).

Authors' Contributions

WZ, YEC, CSF, PT, WKL, PK conceived the study. WKL and PK supervised the research. PT, KKY, WKL, and PK acquired funding. JXT, SD, WH, JY, SC, PT, CWC, KKY, WKL performed data acquisition and data curation. WZ, YEC, CSF, JZ, PK developed the analysis methodology. WZ, YEC, JZ wrote software, performed data analysis and visualization. WZ and PK led the manuscript writing, with critical inputs from YEC, CSF, and WKL. All authors interpreted the findings, reviewed, and approved the final manuscript. WKL and PKS are the corresponding authors of this study, and can be reached by email at wengkhong.lim@duke-nus.edu.sg and pavitrak@i2r.a-star.edu.sg respectively.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary information.
[PDF File (Adobe PDF File), 1384 KB - jmir_v24i7e34669_app1.pdf]

Multimedia Appendix 2
Supplementary data: code for feature generation.
[ZIP File (Zip Archive), 18 KB - jmir_v24i7e34669_app2.zip]

References

1. Vogels EA. About one-in-five Americans use a smart watch or fitness tracker. Pew Research Center. 2020 Jan 9. URL: https://www.pewresearch.org/fact-tank/2020/01/09/about-one-in-five-americans-use-a-smart-watch-or-fitness-tracker/ [accessed 2021-09-07]
2. Li X, Dunn J, Salins D, Zhou G, Zhou W, Schüssler-Fiorenza Rose SM, et al. Digital health: tracking physiomes and activity using wearable biosensors reveals useful health-related information. PLoS Biol 2017 Jan 12;15(1):e2001402 [FREE Full text] [doi: 10.1371/journal.pbio.2001402] [Medline: 28081144]
3. Mishra T, Wang M, Metwally AA, Bogu GK, Brooks AW, Bahmani A, et al. Pre-symptomatic detection of COVID-19 from smartwatch data. Nat Biomed Eng 2020 Dec;4(12):1208-1220 [FREE Full text] [doi: 10.1038/s41551-020-00640-6] [Medline: 33208926]
4. Lim WK, Davila S, Teo JX, Yang C, Pua CJ, Blöcker C, et al. Beyond fitness tracking: the use of consumer-grade wearable data from normal volunteers in cardiovascular and lipidomics research. PLoS Biol 2018 Feb 27;16(2):e2004285 [FREE Full text] [doi: 10.1371/journal.pbio.2004285] [Medline: 29485983]
5. Quisel T, Foschini L, Kale DC. Intra-day activity better predicts chronic conditions. In: Workshop on Machine Learning for Health. 2016 Presented at: NeurIPS '16; December 5-10, 2016; Barcelona, Spain URL: http://arxiv.org/abs/1612.01200
6. 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;18(8):581-599 [FREE Full text] [doi: 10.1038/s41569-021-00522-7] [Medline: 33664502]
7. Garcia-Ceja E, Riegler M, Nordgreen T, Jakobsen P, Oedegaard KJ, Tørresen J. Mental health monitoring with multimodal sensing and machine learning: a survey. Pervasive Mob Comput 2018 Dec;51:1-26. [doi: 10.1016/j.pmcj.2018.09.003]
8. Cooney MT, Vartiainen E, Laatikainen T, Juolevi A, Dudina A, Graham IM. Elevated resting heart rate is an independent risk factor for cardiovascular disease in healthy men and women. Am Heart J 2010 Apr;159(4):612-9.e3. [doi: 10.1016/j.ahj.2009.12.029] [Medline: 20362720]

9. Fox K, Borer JS, Camm AJ, Danchin N, Ferrari R, Lopez Sendon JL, Heart Rate Working Group. Resting heart rate in cardiovascular disease. J Am Coll Cardiol 2007 Aug 28;50(9):823-830 [FREE Full text] [doi: 10.1016/j.jacc.2007.04.079] [Medline: 17719466]
10. Cook S, Togni M, Schaub MC, Wenaweser P, Hess OM. High heart rate: a cardiovascular risk factor? Eur Heart J 2006 Oct;27(20):2387-2393. [doi: 10.1093/eurheartj/ehl259] [Medline: 17000632]
11. Rykov Y, Thach T, Dunleavy G, Roberts AC, Christopoulos G, Soh CK, et al. Activity tracker-based metrics as digital markers of cardiometabolic health in working adults: cross-sectional study. JMIR Mhealth Uhealth 2020 Jan 31;8(1):e16409 [FREE Full text] [doi: 10.2196/16409] [Medline: 32012098]
12. Bumgarner JM, Lambert CT, Hussein AA, Cantillon DJ, Baranowski B, Wolski K, et al. Smartwatch algorithm for automated detection of atrial fibrillation. J Am Coll Cardiol 2018 May 29;71(21):2381-2388 [FREE Full text] [doi: 10.1016/j.jacc.2018.03.003] [Medline: 29535065]
13. Tison GH, Sanchez JM, Ballinger B, Singh A, Olgin JE, Pletcher MJ, et al. Passive detection of atrial fibrillation using a commercially available smartwatch. JAMA Cardiol 2018 May 01;3(5):409-416 [FREE Full text] [doi: 10.1001/jamacardio.2018.0136] [Medline: 29562087]
14. Turakhia MP, Hoang DD, Zimetbaum P, Miller JD, Froelicher VF, Kumar UN, et al. Diagnostic utility of a novel leadless arrhythmia monitoring device. Am J Cardiol 2013 Aug 15;112(4):520-524 [FREE Full text] [doi: 10.1016/j.amjcard.2013.04.017] [Medline: 23672988]
15. Galloway CD, Valys AV, Petterson FL, Gundotra VP, Treiman DL, Albert DE, et al. Non-invasive detection of hyperkalemia with a smartphone electrocardiogram and artificial intelligence. J Am Coll Cardiol 2018 Mar;71(11):A272. [doi: 10.1016/s0735-1097(18)30813-1]
16. Galloway CD, Valys AV, Shreibati JB, Treiman DL, Petterson FL, Gundotra VP, et al. Development and validation of a deep-learning model to screen for hyperkalemia from the electrocardiogram. JAMA Cardiol 2019 May 01;4(5):428-436 [FREE Full text] [doi: 10.1001/jamacardio.2019.0640] [Medline: 30942845]
17. Quer G, Gouda P, Galarnyk M, Topol EJ, Steinhubl SR. Inter- and intraindividual variability in daily resting heart rate and its associations with age, sex, sleep, BMI, and time of year: retrospective, longitudinal cohort study of 92,457 adults. PLoS One 2020 Feb 5;15(2):e0227709 [FREE Full text] [doi: 10.1371/journal.pone.0227709] [Medline: 32023264]
18. Sopic D, Aminifar A, Aminifar A, Atienza D. Real-time event-driven classification technique for early detection and prevention of myocardial infarction on wearable systems. IEEE Trans Biomed Circuits Syst 2018 Jul 16;12(5):982-992. [doi: 10.1109/TBCAS.2018.2848477] [Medline: 30010598]
19. Attia ZI, Kapa S, Lopez-Jimenez F, McKie PM, Ladewig DJ, Satam G, et al. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. Nat Med 2019 Jan;25(1):70-74. [doi: 10.1038/s41591-018-0240-2] [Medline: 30617318]
20. Strath SJ, Rowley TW. Wearables for promoting physical activity. Clin Chem 2018 Jan;64(1):53-63. [doi: 10.1373/clinchem.2017.272369] [Medline: 29118062]
21. Teo JX, Davila S, Yang C, Hii AA, Pua CJ, Yap J, et al. Digital phenotyping by consumer wearables identifies sleep-associated markers of cardiovascular disease risk and biological aging. Commun Biol 2019 Oct 4;2:361 [FREE Full text] [doi: 10.1038/s42003-019-0605-1] [Medline: 31602410]
22. Guo C, Lu M, Chen J. An evaluation of time series summary statistics as features for clinical prediction tasks. BMC Med Inform Decis Mak 2020 Mar 05;20(1):48 [FREE Full text] [doi: 10.1186/s12911-020-1063-x] [Medline: 32138733]
23. Schroeder EB, Liao D, Chambless LE, Prineas RJ, Evans GW, Heiss G. Hypertension, blood pressure, and heart rate variability: the Atherosclerosis Risk in Communities (ARIC) study. Hypertension 2003 Dec;42(6):1106-1111. [doi: 10.1161/01.HYP.0000100444.71069.73] [Medline: 14581296]
24. Liao D, Carnethon M, Evans GW, Cascio WE, Heiss G. Lower heart rate variability is associated with the development of coronary heart disease in individuals with diabetes: the atherosclerosis risk in communities (ARIC) study. Diabetes 2002 Dec;51(12):3524-3531. [doi: 10.2337/diabetes.51.12.3524] [Medline: 12453910]
25. Kotecha D, New G, Flather MD, Eccleston D, Krum H. 61 Five-min heart rate variability can predict obstructive angiographic coronary disease. Heart 2011 Jun 09;97(Suppl 1):A38-A39. [doi: 10.1136/heartjnl-2011-300198.61]
26. Harvard Health Publishing Staff. Heart rate variability: how it might indicate well-being. Harvard Health. 2021 Dec 1. URL: https://www.health.harvard.edu/blog/heart-rate-variability-new-way-track-well-2017112212789 [accessed 2021-09-07]
27. Oura Team. Heart Rate During Sleep: Look for These 3 Patterns. Oura. 2020 Feb 11. URL: https://ouraring.com/blog/heart-rate-during-sleep/ [accessed 2021-09-07]
28. Pevnick JM, Birkeland K, Zimmer R, Elad Y, Kedan I. Wearable technology for cardiology: an update and framework for the future. Trends Cardiovasc Med 2018 Feb;28(2):144-150 [FREE Full text] [doi: 10.1016/j.tcm.2017.08.003] [Medline: 28818431]
29. Nelson BW, Allen NB. Accuracy of consumer wearable heart rate measurement during an ecologically valid 24-hour period: intraindividual validation study. JMIR Mhealth Uhealth 2019 Mar 11;7(3):e10828 [FREE Full text] [doi: 10.2196/10828] [Medline: 30855232]

30. Godino JG, Wing D, de Zambotti M, Baker FC, Bagot K, Inkelis S, et al. Performance of a commercial multi-sensor wearable (Fitbit Charge HR) in measuring physical activity and sleep in healthy children. PLoS One 2020 Sep 4;15(9):e0237719 [FREE Full text] [doi: 10.1371/journal.pone.0237719] [Medline: 32886714]
31. Ballinger B, Hsieh J, Singh A, Sohoni N, Wang J, Tison GH, et al. DeepHeart: semi-supervised sequence learning for cardiovascular risk prediction. Proc AAAI Conf Artif Intell 2018 Apr 26;32(1):2079-2086 [FREE Full text] [doi: 10.1609/aaai.v32i1.11891]
32. Tison GH, Singh AC, Ohashi DA, Hsieh JT, Ballinger BM, Olgin JE, et al. Abstract 21042: cardiovascular risk stratification using off-the-shelf wearables and a multi-task deep learning algorithm. Circulation 2017 Nov 14;136(suppl_1):A21042. [doi: 10.1161/circ.136.suppl_1.21042]
33. National Heart Centre Singapore. Effects of physical activity, ambulatory blood pressure and calcium score on cardiovascular health in normal people (SingHEART) – Report No.: NCT02791152. Clinical Trials. 2016. URL: https://clinicaltrials.gov/ct2/show/NCT02791152 [accessed 2021-11-01]
34. Yap J, Lim WK, Sahlén A, Chin CW, Chew KM, Davila S, et al. Harnessing technology and molecular analysis to understand the development of cardiovascular diseases in Asia: a prospective cohort study (SingHEART). BMC Cardiovasc Disord 2019 Nov 21;19(1):259 [FREE Full text] [doi: 10.1186/s12872-019-1248-3] [Medline: 31752689]
35. Fulcher BD, Jones NS. Highly comparative feature-based time-series classification. IEEE Trans Knowl Data Eng 2014 Dec 1;26(12):3026-3037. [doi: 10.1109/tkde.2014.2316504]
36. Fulcher BD, Jones NS. hctsa: a computational framework for automated time-series phenotyping using massive feature extraction. Cell Syst 2017 Nov 22;5(5):527-31.e3 [FREE Full text] [doi: 10.1016/j.cels.2017.10.001] [Medline: 29102608]
37. Christ M, Braun N, Neuffer J, Kempa-Liehr AW. Time series FeatuRe extraction on basis of scalable hypothesis tests (tsfresh – a Python package). Neurocomputing 2018 Sep 13;307:72-77. [doi: 10.1016/j.neucom.2018.03.067]
38. Christ M, Kempa-Liehr AW, Feindt M. Distributed and parallel time series feature extraction for industrial big data applications. arXiv 2017 May 19 [FREE Full text]
39. Lubba CH, Sethi SS, Knaute P, Schultz SR, Fulcher BD, Jones NS. catch22: CAnonical Time-series CHaracteristics. Data Min Knowl Disc 2019 Aug 09;33(6):1821-1852. [doi: 10.1007/s10618-019-00647-x]
40. Bent B, Goldstein BA, Kibbe WA, Dunn JP. Investigating sources of inaccuracy in wearable optical heart rate sensors. NPJ Digit Med 2020 Feb 10;3:18 [FREE Full text] [doi: 10.1038/s41746-020-0226-6] [Medline: 32047863]
41. Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H. UpSet: visualization of intersecting sets. IEEE Trans Vis Comput Graph 2014 Dec;20(12):1983-1992 [FREE Full text] [doi: 10.1109/TVCG.2014.2346248] [Medline: 26356912]
42. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics 2017 Sep 15;33(18):2938-2940 [FREE Full text] [doi: 10.1093/bioinformatics/btx364] [Medline: 28645171]
43. Breiman L. Random forests. Mach Learn 2001 Oct 1;45(1):5-32. [doi: 10.1023/A:1010933404324]
44. Liaw A, Wiener M. Classification and regression by RandomForest. R News 2002 Dec;2(3):18-22.
45. Chen C, Liaw A, Breiman L. Using Random Forest to Learn Imbalanced Data. Department of Statistics, University of California, Berkeley. 2004 Jul 1. URL: https://statistics.berkeley.edu/tech-reports/666 [accessed 2021-09-07]
46. Brier GW. Verification of forecasts expressed in terms of probability. Mon Wea Rev 1950 Jan 01;78(1):1-3. [doi: 10.1175/1520-0493(1950)078<0001:vofeit>2.0.co;2]
47. Breiman L. Out-of-bag estimation. University of California Berkeley. 1996. URL: https://www.stat.berkeley.edu/~breiman/OOBestimation.pdf [accessed 2021-11-01]
48. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Methodol 1995;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]
49. The Polygenic Score (PGS) Catalog. PGS Catalog. URL: https://www.pgscatalog.org/ [accessed 2021-09-07]
50. Lundberg S, Lee SI. An unexpected unity among methods for interpreting model predictions. In: Advances in Neural Information Processing Systems. 2016 Dec Presented at: NeurIPS '16; December 5-10, 2016; Barcelona, Spain URL: http://arxiv.org/abs/1611.07478
51. Molnar C, Casalicchio G, Bischl B. iml: an R package for interpretable machine learning. J Open Source Softw 2018 Jun 27;3(27):786. [doi: 10.21105/joss.00786]
52. Pouromran F, Radhakrishnan S, Kamarthi S. Exploration of physiological sensors, features, and machine learning models for pain intensity estimation. PLoS One 2021 Jul 9;16(7):e0254108 [FREE Full text] [doi: 10.1371/journal.pone.0254108] [Medline: 34242325]
53. Hu B, Chen Y, Keogh E. Classification of streaming time series under more realistic assumptions. Data Min Knowl Disc 2015 Jun 3;30(2):403-437. [doi: 10.1007/s10618-015-0415-0]
54. Kubota Y, Chen LY, Whitsel EA, Folsom AR. Heart rate variability and lifetime risk of cardiovascular disease: the Atherosclerosis Risk in Communities Study. Ann Epidemiol 2017 Oct;27(10):619-25.e2 [FREE Full text] [doi: 10.1016/j.annepidem.2017.08.024] [Medline: 29033120]
55. van de Vegte YJ, van der Harst P, Verweij N. Heart rate recovery 10 seconds after cessation of exercise predicts death. J Am Heart Assoc 2018 Apr 05;7(8):e008341 [FREE Full text] [doi: 10.1161/JAHA.117.008341] [Medline: 29622586]

56. Georgiopoulou VV, Kalogeropoulos AP, Chowdhury R, Binongo JN, Bibbins-Domingo K, Rodondi N, Health ABC Study. Exercise capacity, heart failure risk, and mortality in older adults: the health ABC study. Am J Prev Med 2017 Feb;52(2):144-153 [FREE Full text] [doi: 10.1016/j.amepre.2016.08.041] [Medline: 27856115]
57. Niemelä M, Kangas M, Farrahi V, Kiviniemi A, Leinonen AM, Ahola R, et al. Intensity and temporal patterns of physical activity and cardiovascular disease risk in midlife. Prev Med 2019 Jul;124:33-41 [FREE Full text] [doi: 10.1016/j.ypmed.2019.04.023] [Medline: 31051183]
58. Khera AV, Emdin CA, Drake I, Natarajan P, Bick AG, Cook NR, et al. Genetic risk, adherence to a healthy lifestyle, and coronary disease. N Engl J Med 2016 Dec 15;375(24):2349-2358 [FREE Full text] [doi: 10.1056/NEJMoa1605086] [Medline: 27959714]
59. Said MA, Verweij N, van der Harst P. Associations of combined genetic and lifestyle risks with incident cardiovascular disease and diabetes in the UK Biobank study. JAMA Cardiol 2018 Aug 01;3(8):693-702 [FREE Full text] [doi: 10.1001/jamacardio.2018.1717] [Medline: 29955826]
60. Individual access to genomic disease risk factors has a beneficial impact on lifestyles. ScienceDaily. 2018 Jun 15. URL: https://www.sciencedaily.com/releases/2018/06/180615185408.htm [accessed 2021-09-07]
61. Ottman R. Gene-environment interaction: definitions and study designs. Prev Med 1996;25(6):764-770 [FREE Full text] [doi: 10.1006/pmed.1996.0117] [Medline: 8936580]
62. Thomas D. Methods for investigating gene-environment interactions in candidate pathway and genome-wide association studies. Annu Rev Public Health 2010;31:21-36 [FREE Full text] [doi: 10.1146/annurev.publhealth.012809.103619] [Medline: 20070199]
63. Li J, Li X, Zhang S, Snyder M. Gene-environment interaction in the era of precision medicine. Cell 2019 Mar 21;177(1):38-44 [FREE Full text] [doi: 10.1016/j.cell.2019.03.004] [Medline: 30901546]
64. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 2015 Mar 31;12(3):e1001779 [FREE Full text] [doi: 10.1371/journal.pmed.1001779] [Medline: 25826379]
65. Muggeridge DJ, Hickson K, Davies AV, Giggins OM, Megson IL, Gorely T, et al. Measurement of heart rate using the polar OH1 and Fitbit charge 3 wearable devices in healthy adults during light, moderate, vigorous, and sprint-based exercise: validation study. JMIR Mhealth Uhealth 2021 Mar 25;9(3):e25313 [FREE Full text] [doi: 10.2196/25313] [Medline: 33764310]
66. Lubitz SA, Faranesh AZ, Atlas SJ, McManus DD, Singer DE, Pagoto S, et al. Rationale and design of a large population study to validate software for the assessment of atrial fibrillation from data acquired by a consumer tracker or smartwatch: the Fitbit heart study. Am Heart J 2021 Aug;238:16-26. [doi: 10.1016/j.ahj.2021.04.003] [Medline: 33865810]
67. National Supercomputing Centre. URL: https://www.nscc.sg/ [accessed 2022-07-04]

Abbreviations

Catch22: Canonical Time-series Characteristics 22
MDA: mean decrease in accuracy
OOB: out-of-bag
PGS: polygenic score
SHAP: Shapley Additive Explanations

Edited by R Kukafka; submitted 09.11.21; peer-reviewed by G Chevance; comments to author 04.12.21; revised version received 12.04.22; accepted 29.05.22; published 29.07.22.

Please cite as:
Zhou W, Chan YE, Foo CS, Zhang J, Teo JX, Davila S, Huang W, Yap J, Cook S, Tan P, Chin CWL, Yeo KK, Lim WK, Krishnaswamy P
High-Resolution Digital Phenotypes From Consumer Wearables and Their Applications in Machine Learning of Cardiometabolic Risk Markers: Cohort Study
J Med Internet Res 2022;24(7):e34669
URL: https://www.jmir.org/2022/7/e34669
doi:10.2196/34669
PMID:35904853

Original Paper

Patient-Reported Outcome and Experience Measures in Perinatal Care to Guide Clinical Practice: Prospective Observational Study

Anne Louise Depla[1], MD; Marije Lamain-de Ruiter[1,2], MD, PhD; Lyzette T Laureij[2], MD; Hiske E Ernst-Smelt[2], Drs; Jan A Hazelzet[3], MD, PhD; Arie Franx[2], MD, PhD; Mireille N Bekker[1], MD, PhD[‡]; BUZZ Project Team[4]

[1]Department of Obstetrics and Gynecology, Wilhemina Children's Hospital, University Medical Center Utrecht, Utrecht, Netherlands

[2]Department of Obstetrics and Gynecology, Division of Obstetrics and Fetal Medicine, Erasmus Medical Center Sophia, Rotterdam, Netherlands

[3]Department of Public Health, Erasmus Medical Center, University Medical Center Rotterdam, Rotterdam, Netherlands

[4]See Acknowledgments

[‡]Written on behalf of the BUZZ project team

Corresponding Author:
Mireille N Bekker, MD, PhD
Department of Obstetrics and Gynecology
Wilhemina Children's Hospital
University Medical Center Utrecht
Lundlaan 6
Utrecht, 3584 EA
Netherlands
Phone: 31 88 7553981
Email: M.N.Bekker-3@umcutrecht.nl

Abstract

Background: The International Consortium for Health Outcomes Measurement has published a set of patient-centered outcome measures for pregnancy and childbirth (PCB set), including patient-reported outcome measures (PROMs) and patient-reported experience measures (PREMs). To establish value-based pregnancy and childbirth care, the PCB set was implemented in the Netherlands, using the outcomes on the patient level for shared decision-making and on an aggregated level for quality improvement.

Objective: This study aims to report first outcomes, experiences, and practice insights of implementing the PCB set in clinical practice.

Methods: In total, 7 obstetric care networks across the Netherlands, each consisting of 1 or 2 hospitals and multiple community midwifery practices (ranging in number from 2 to 18), implemented the PROM and PREM domains of the PCB set as part of clinical routine. This observational study included all women participating in the clinical project. PROMs and PREMs were assessed with questionnaires at 5 time points: 2 during pregnancy and 3 post partum. Clinical threshold values (alerts) supported care professionals interpreting the answers, indicating possibly alarming outcomes per domain. Data collection took place from February 2020 to September 2021. Data analysis included missing (pattern) analysis, sum scores, alert rates, and sensitivity analysis.

Results: In total, 1923 questionnaires were collected across the 5 time points: 816 (42.43%) at T1 (first trimester), 793 (41.23%) at T2 (early third trimester), 125 (6.5%) at T3 (maternity week), 170 (8.84%) at T4 (6 weeks post partum), and 19 (1%) at T5 (6 months post partum). Of these, 84% (1615/1923) were filled out completely. Missing items per domain ranged from 0% to 13%, with the highest missing rates for depression, pain with intercourse, and experience with pain relief at birth. No notable missing patterns were found. For the PROM domains, relatively high alert rates were found both in pregnancy and post partum for incontinence (469/1798, 26.08%), pain with intercourse (229/1005, 22.79%), breastfeeding self-efficacy (175/765, 22.88%), and mother-child bonding (122/288, 42.36%). Regarding the PREM domains, the highest alert rates were found for birth experience (37/170, 21.76%), shared decision-making (101/982, 10.29%), and discussing pain relief ante partum (310/793, 39.09%). Some domains showed very little clinical variation; for example, role of the mother and satisfaction with care.

Conclusions: The PCB set is a useful tool to assess patient-reported outcomes and experiences that need to be addressed over the whole course of pregnancy and childbirth. Our results provide opportunities to improve and personalize perinatal care. Furthermore, we could propose several recommendations regarding methods and timeline of measurements based on our findings.

This study supports the implementation of the PCB set in clinical practice, thereby advancing the transformation toward patient-centered, value-based health care for pregnancy and childbirth.

(J Med Internet Res 2022;24(7):e37725) doi:10.2196/37725

KEYWORDS

perinatal care; patient-reported outcomes; patient-reported experiences; patient-centered outcome measures; value-based health care; shared decision-making; personalized care; integrated care

Introduction

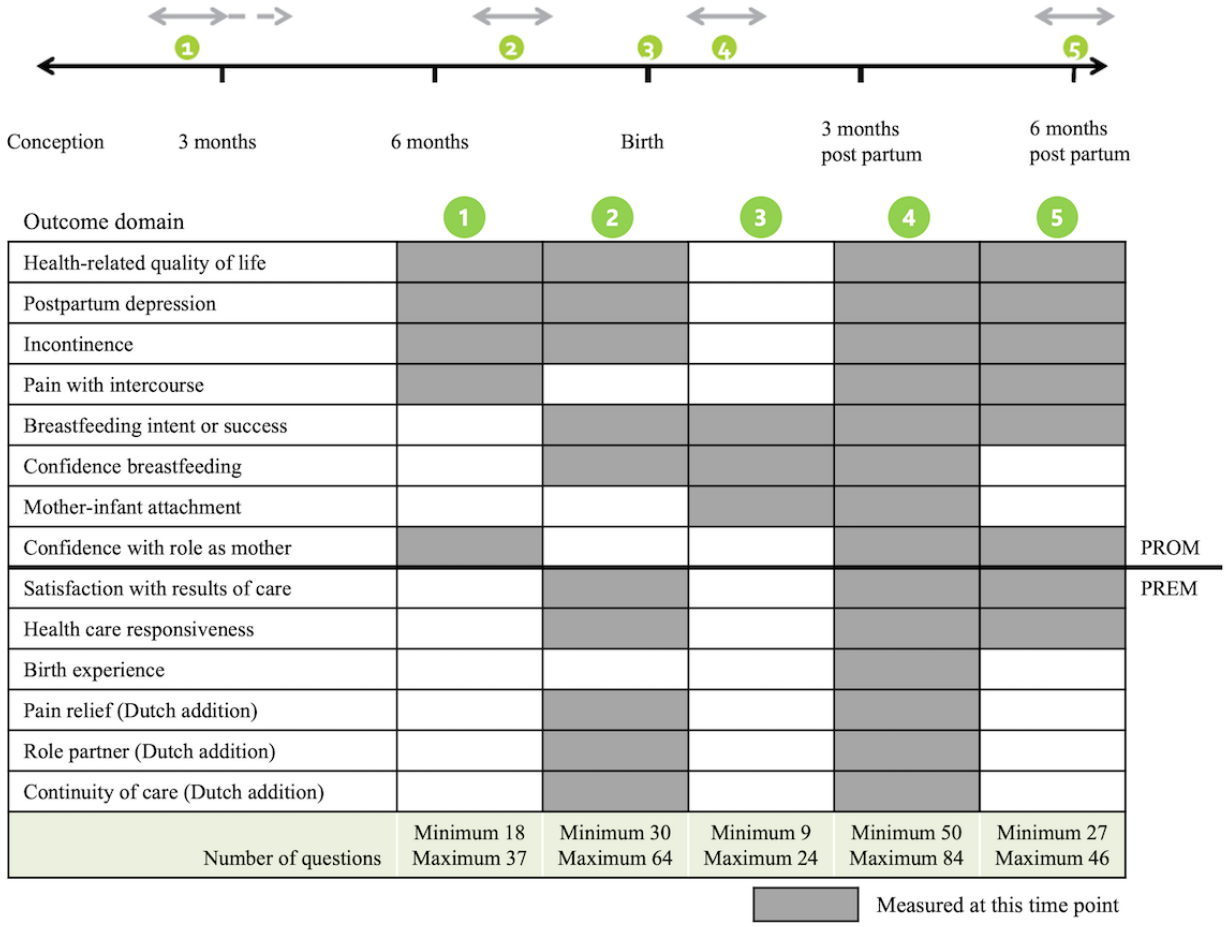
Background

Currently, health care systems are moving toward high-value care, adapted to each individual patient [1,2]. These health care systems prioritize patients' health goals in care decisions and quality improvement, above processes and clinical parameters. The transformation into a patient-centered, value-driven system is dependent on access to data that capture what matters most to patients [3-5]. Patient-reported outcome measures (PROMs) and patient-reported experience measures (PREMs) provide standardized assessment of patients' health status or experience with health care directly from the patient [6]. Integrated into routine care, these measures can facilitate patient-provider communication, improve patients' experiences, and enhance detection and management of their health status [7-9]. When aggregated, PROMs and PREMs foster inclusion of patients' perspective in continuous quality improvement, along with clinical measures that have already been captured for quality performance [10].

Just as in other disciplines, perinatal care may benefit from systematic PROM and PREM assessment to enhance quality of care. Moreover, patient-reported outcomes of perinatal care, such as depression or incontinence, may have serious long-term consequences for the health of the mother and child and might currently be undervalued. The interest in, and use of, PROMs and PREMs has grown in perinatal care, but most PROMs and PREMs in this field are assessed anonymously for quality improvement or research purposes only [11], whereas PROMs and PREMs, if integrated in clinical care on an individual level, could provide perinatal caregivers an opportunity to detect symptoms and adapt care appropriately, as well as encourage patients to think, and speak, about their current well-being and experiences [12]. Nevertheless, clinical integration of PROMs and PREMs has many challenges such as selecting relevant topics, valid assessment instruments, measurement moments, and threshold values that require action [3,13,14].

The International Consortium for Health Outcomes Measurement (ICHOM) has published a core set of patient-centered outcome measures for pregnancy and childbirth (PCB set), proposing standardized measures of clinical outcomes as well as patient outcomes and experiences over the full cycle of care [15]. For its patient-reported domains, the PCB set includes measurement instruments (ie, questionnaires) and a timeline for assessment: at 5 time points throughout pregnancy and post partum until 6 months after birth (Figure 1 [16]). Recently, the feasibility and acceptability of the PCB set were studied in clinic and its patient-reported domains collected for research purposes [17-19]. In addition, some of its measurement instruments were evaluated for validity and reliability in a maternity population [20-22]. However, little is known regarding compliance with the PROM and PREM questionnaires of the PCB set and the clinical performance of threshold values that require action throughout pregnancy and the postpartum period.

Figure 1. Pregnancy and childbirth outcome set: patient-reported domains and moments to measure (adapted from Nijagal et al [16]). PREM: patient-reported experience measure; PROM: patient-reported outcome measure.



Study Rationale

In an implementation project across the Netherlands, 7 regions incorporated the PCB set in clinic over the full cycle of perinatal care with all care professionals involved. In the journey toward value-based perinatal care, the primary goal was to discuss individual PROMs and PREMs as part of regular care and use them for shared decision-making to personalize care accordingly (level 1 of value-based health care). Furthermore, aggregated PROM and PREM results could be used for patient-centered quality improvement (level 2 of value-based health care). During the project, we closely monitored first experiences and practice insights of the regions' incorporation of patient-reported measures into routine perinatal care at an individual level. This study aimed to report compliance with the PROM and PREM questionnaires, the outcomes per domain throughout pregnancy and post partum, and the clinical use of threshold values. Our findings can support clinical implementation of value-based health care with the PCB set, accelerate the transformation toward personalized care, and contribute to governance of the PCB set to retain its international comparability.

Methods

Study Design

An observational study was conducted to report and gain insight into PROMs and PREMs as part of clinical routine for personalized perinatal care. This paper is written following the Strengthening the Reporting of Observational Studies in Epidemiology checklist [23].

Setting

This study was carried out as part of a project involving the implementation of the PCB set in Dutch perinatal care called the Dutch abbreviation of Discuss Outcomes of Pregnancy with the Pregnant Woman (BUZZ) project. In total, 7 regions across the Netherlands joined forces to implement the PROM and PREM domains of the PCB set in routine clinical practice. The implementation was supported by Zorginstituut Nederland and coincided with a nationwide ministry program to enhance value-based health care and shared decision-making [24]. Each participating region consisted of 1 or 2 hospitals and 2 to 18 community midwifery practices (Table 1) collaborating in local obstetric care networks (OCNs; refer to Textbox 1 for an explanation of Dutch perinatal care organization). Data were collected from February 2020 to September 2021.

Table 1. Implementation strategy per obstetric care network.

	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7
Time point 1: first trimester			✓	✓	✓	✓	
Time point 2: early third trimester	✓	✓	✓	✓	✓	✓	
Time point 3: maternity week			✓	✓	✓	✓	✓
Time point 4: 6 weeks post partum		✓	✓	✓	✓	✓	✓
Time point 5: 6 months post partum			[a]		✓		✓
Collection	Stand-alone data capture tool	EHR[b]	EHR	Stand-alone data capture tool	Stand-alone data capture tool	Stand-alone data capture tool	Paper
Hospitals	1	1	1	1	1	1	2
Community midwifery practices	3	2	13	2	2	9	18
Patient group	All	All	All	Women in vulnerable situations	Diabetes or history of CS[c]	GBS+[d]	Induction with AROM[e] by CM[f]

[a]Planned to implement at the end of the project period.

[b]EHR: electronic health record.

[c]CS: cesarean section.

[d]GBS+: urine sample positive for Group B streptococcus in pregnancy.

[e]AROM: artificial rupture of membranes.

[f]CM: community midwife.

Textbox 1. Organization of Dutch perinatal care.

Organization of Dutch perinatal care

- Dutch perinatal care is organized in a 2-tier system.
- Community midwives provide primary care for low-risk pregnancies and act as gatekeepers to specialist care. These midwives have their own professional autonomy, responsibilities, and financial arrangements.
- For medium- to high-risk pregnancies, hospital-employed obstetric care professionals provide secondary or tertiary specialist care.
- Of all women receiving perinatal care, up to 70% visit both health care tiers [25].
- Over the last decade, a more integrated obstetric care system has been advised by the ministry of health, which is partly being realized by collaboration of both tiers in obstetric care networks.

Participants

Women receiving perinatal care at a participating organization were invited to complete PROM and PREM questionnaires as part of usual care. Women who additionally gave informed consent to use their answers for research were included in this study. Informed consent was obtained in the PROM and PREM questionnaire itself. As this study aimed to report outcomes of the PCB set as is, we report the results of all PROM and PREM questionnaires collected within the project period; no target size was predetermined.

Implementation in Clinical Practice

The primary purpose of the BUZZ project was to use PROM and PREM questionnaires to guide individual perinatal care. Pregnant and postpartum women were invited to fill out questionnaires as part of routine care and their obstetric care professional discussed the answers in their next regular visit. The BUZZ project was explicitly organized within OCNs to ensure continuity of care over the full cycle of care for pregnancy and childbirth. The project team of each OCN made local decisions to enhance implementation in their practice on several key points (Table 1):

- Mode of administering questionnaires: some sites could capture questionnaires through their electronic health record (EHR), others used a stand-alone data capture tool, and 1 site used paper questionnaires (whatever at that moment was considered the most optimal to use the responses in their clinical setting).

- Population and time points: most sites chose to start small by either selecting a few time points for PROM and PREM assessment or a specific patient group.
- Site-specific adaptations: some sites made minor adaptations to the questionnaire content. For example, 1 site dismissed the screening questions for depression and used the full questionnaire in all women.

Outcome Measures

The PCB set's PROM and PREM domains were captured as proposed by ICHOM with questionnaires at 5 time points during pregnancy and post partum (Figure 1) [16]. Each domain is assessed with its own measurement instrument, consisting of one or more questions (Multimedia Appendix 1). At every time point only relevant domains are assessed. In some domains, one or more screening questions can either rule in or rule out further questions for that domain. To fit Dutch perinatal care, a few domains have been added to the original PCB set (Figure 1) [17]. Before implementation, the translated Dutch questionnaires were tested among 4 women with low health literacy by the Dutch center of expertise on health disparities (Pharos). Minor adaptations were carried out where possible; questionnaires already validated in Dutch were not adapted. For each measurement instrument a clinical threshold value (alert) was defined according to existing literature or, if not available, determined by the multidisciplinary national BUZZ project team, informed by expert opinion (Multimedia Appendix 1). The alerts supported care professionals interpreting the answers, indicating worrisome outcomes through a color-coded dashboard (or calculated by hand in case of paper questionnaires). As clinical data could not yet be merged (digitally), a few casemix variables were collected through the questionnaires: age, gravidity, parity, postal code, and ethnicity.

Data Analysis

Only the data of women who gave informed consent were uploaded by project leaders to a central and highly secure digital research environment. Data merging and analysis was performed on this secured server using R software (version 4.0.2; The R Foundation for Statistical Computing) [26]. Duplicate and blank questionnaires resulting from technical problems were removed. In addition, questionnaires with only the first item filled out, requesting informed consent or social support, were excluded because we could not determine whether this resulted from a technical problem. A new option to answer a question was added by 1 site (ie, *not applicable*): these answers have been considered missing in analysis because they were not included in the national (validated) scoring systems. Secondary analysis of these data was considered, but the numbers were too small. Questions that were answered unintentionally, for example, a full depression questionnaire filled out despite having scored a negative screening, were removed. The casemix variables gravidity and parity are reported as state in current pregnancy: if parity and gravidity were equal, parity was corrected to gravidity–1. Completion rates were calculated per question and per measurement instrument. If applicable, sum scores were calculated according to a predefined scoring system. Missing items were excluded from this calculation; therefore, sum scores with one or more missing items are lower by definition. Alerts were calculated according to the thresholds provided in Multimedia Appendix 1. In an additional sensitivity analysis of domains with multiple questions, results with >25% missing items were removed, and their mean sum scores and alert rates were compared with the complete analysis.

Ethics Approval

The Medical Ethics Review Committee of the Erasmus Medical Center (MEC-2020-0129) declared that the Medical Research Involving Human Subjects Act (WMO) does not apply to this study. Therefore, it was exempt from formal medical ethics assessment. For each site, local approval was obtained from the regional ethics board.

Results

Overall

In total, 1923 unique questionnaires were collected, most of them during pregnancy (Table 2). The median moments of completion corresponded well with the proposed time points (Figure 1). Some T2 and T4 questionnaires were completed earlier than the proposed window, whereas a few T1 questionnaires were filled out too late. The questionnaires were filled out by 1318 individual women, of whom 838 (63.58%) completed 1 questionnaire, and the remaining 480 (36.41%) completed up to 4 questionnaires. Their baseline characteristics are presented in Table 3. Sum scores and alerts per domain and time point are presented in Tables 4 and 5. Multimedia Appendix 2 contains figures that show each domain's scores and alerts.

Table 2. Moments of questionnaire completion (N=1923).

Time point	Value, n (%)	Moment of questionnaire completion, median (range)
First trimester (T1)	816 (42.43)	15 (9-27)[a]
Early third trimester (T2)	793 (41.24)	28 (23-37)[a]
Maternity week[b] (T3)	125 (6.5)	5 (4-5)[c]
Post partum, 6 weeks[b] (T4)	170 (8.84)	3 (0-12)[d]
Post partum, 6 months (T5)	19 (1)	27 (22-30)[d]

[a]Moment occurred in weeks of pregnancy.

[b]The exact moment of completion was missing for *maternity week* and *6 weeks post partum* for 123 and 127 questionnaires, respectively. Because of the information technology system setup, we do know that *maternity week* questionnaires were completed mostly between 1 and 3 weeks post partum and *6 weeks post partum* questionnaires between 3 and 5 weeks post partum.

[c]Moment occurred in days post partum.

[d]Moment occurred in weeks post partum.

Table 3. Participant characteristics (N=1318).

Characteristics	Values
Age (years), median (range); missing: n=77	32 (17-46)
Parity, n (%); missing: n=330	
Nulliparous	360 (36.43)
Multiparous	628 (63.56)
Ethnicity, n (%); missing: n=143	
Western	1057 (89.96)
Other	118 (10.04)

Table 4. Outcomes per patient-reported outcome measure domain.

Domain and subdomain	Time point	Value, n (%)	Score, median (range)	Alerts, n (%)	Missing[a], n (%)
Social support	All	1092 (56.79)	3 (0-3)	44 (4.06)	7 (0.64)
Quality of life					
	All	1798 (93.5)	37 (7-50)	21 (1.17)	1 (0.06)
	T1[b]	816 (45.38)	38 (7-50)	6 (0.74)	0 (0)
	T2[c]	793 (44.1)	37 (7-50)	12 (1.52)	1 (0.13)
	T4[d]	170 (9.45)	38 (14-49)	2 (1.18)	0 (0)
	T5[e]	19 (1.06)	37 (19-46)	1 (5.26)	0 (0)
Mental health					
Screen depression					
	All	1756 (91.32)	0 (0-6)	61 (3.52)	25 (1.42)
	T1	798 (45.44)	0 (0-6)	33 (4.19)	10 (1.25)
	T2	776 (44.19)	0 (0-5)	22 (2.85)	5 (0.64)
	T4	163 (9.28)	0 (0-5)	5 (3.27)	10 (6.13)
	T5	19 (1.08)	0 (0-4)	1 (5.26)	0 (0)
Full depression[f]					
	All	103 (5.36)	10 (0-25)	47 (52.22)	13 (12.62)
	T1	51 (49.51)	11 (0-23)	27 (52.94)	0 (0)
	T2	39 (37.86)	7 (0-25)	13 (44.83)	10 (25.64)
	T4	12 (11.65)	12 (3-25)	6 (66.67)	3 (25)
	T5	1 (0.97)	N/A[g]	1 (100)	0 (0)
Incontinence and dyspareunia					
Screen, urine					
	All	1798 (93.5)	—[h]	469 (26.91)	55 (3.06)
	T1	816 (45.38)	—	150 (20.15)	22 (2.7)
	T2	793 (44.1)	—	266 (34.64)	25 (3.15)
	T4	170 (9.45)	—	45 (27.78)	8 (4.7)
	T5	19 (1.06)	—	8 (42.1)	0 (0)
Screen, stool					
	All	1798 (93.5)	—	15 (0.86)	57 (3.17)
	T1	816 (45.38)	—	3 (0.38)	23 (2.82)
	T2	793 (44.1)	—	6 (0.78)	26 (3.28)
	T4	170 (9.45)	—	6 (3.70)	8 (4.71)
	T5	19 (1.06)	—	0 (0)	0 (0)
Screen, flatus					
	All	1798 (93.5)	—	388 (22.26)	55 (3.06)
	T1	816 (45.38)	—	149 (18.77)	22 (2.7)
	T2	793 (44.1)	—	190 (24.74)	25 (3.15)
	T4	170 (9.45)	—	44 (27.16)	8 (4.71)
	T5	19 (1.06)	—	5 (26.32)	0 (0)
Full urine[f]					

Domain and subdomain	Time point	Value, n (%)	Score, median (range)	Alerts, n (%)	Missing[a], n (%)
	All	469 (24.39)	6 (0-18)	185 (39.45)	0 (0)
	T1	150 (31.98)	6 (0-15)	62 (41.33)	0 (0)
	T2	266 (56.72)	5 (1-18)	100 (37.59)	0 (0)
	T4	45 (9.59)	6 (1-15)	19 (42.22)	0 (0)
	T5	8 (1.71)	7 (3-12)	4 (50)	0 (0)
Full stool and flatus[f]					
	All	394 (20.49)	3 (0-17)	385 (97.96)	1 (0.25)
	T1	151 (38.32)	3 (0-10)	147 (98)	1 (0.66)
	T2	193 (48.98)	3 (0-14)	190 (98.45)	0 (0)
	T4	45 (11.42)	3 (0-17)	43 (95.56)	0 (0)
	T5	5 (1.27)	2 (2-3)	5 (100)	0 (0)
Pain with intercourse					
	All	1005 (52.26)	0 (0-5)	229 (24.65)	76 (7.56)
	T1	816 (81.19)	0 (0-5)	161 (20.72)	39 (4.78)
	T4	170 (16.91)	1 (0-5)	59 (44.36)	37 (21.76)
	T5	19 (1.89)	0 (0-5)	9 (47.37)	0 (0)
Breastfeeding					
Breastfeeding intention	All (T2)	793 (41.24)	—	172 (22.4)[i]	25 (3.15)
Breastfeeding success					
	All	314 (39.6)	—	116 (39.46)[j]	20 (6.37)
	T3[k]	125 (39.81)	—	45 (36)[j]	0 (0)
	T4	170 (54.14)	—	61 (40.67)[j]	20 (11.76)
	T5	19 (6.05)	—	10 (52.63)[j]	0 (0)
Screen, breastfeeding confidence[f]					
	All	765 (39.78)	4 (1-5)	175 (23)	4 (0.52)
	T2	596 (77.91)	4 (1-5)	150 (25.25)	2 (0.34)
	T3	80 (10.46)	4 (2-5)	13 (16.46)	1 (1.25)
	T4	89 (11.63)	4 (1-5)	12 (13.64)	1 (1.12)
Full breastfeeding self-efficacy[f]					
	All	175 (9.1)	40 (4-64)	124 (72.94)	5 (2.86)
	T2	150 (85.71)	41 (14-64)	104 (71.23)	4 (2.67)
	T3	13 (7.43)	36 (12-54)	11 (84.62)	0 (0)
	T4	12 (6.86)	27 (4-52)	9 (81.82)	1 (8.33)
Role transition					
Mother-child bonding					
	All	288 (14.98)	2 (0-11)	122 (44.85)	16 (5.56)
	T3	125 (43.4)	2 (0-8)	56 (45.9)	3 (2.4)
	T4	163 (56.6)	2 (0-11)	66 (44)	13 (7.98)
Role as mother					
	All	1005 (52.26)	4 (1-5)	3 (0.31)	40 (3.98)
	T1	816 (81.19)	4 (2-5)	1 (0.13)	26 (3.19)

Domain and subdomain	Time point	Value, n (%)	Score, median (range)	Alerts, n (%)	Missing[a], n (%)
	T4	170 (16.91)	5 (2-5)	1 (0.64)	14 (8.24)
	T5	19 (1.89)	5 (1-5)	1 (5.26)	0 (0)

[a]Completely missing.
[b]T1: first trimester.
[c]T2: early third trimester.
[d]T4: 6 weeks post partum.
[e]T5: 6 months post partum.
[f]Optional subdomain, dependent on screening question or questions.
[g]N/A: not applicable.
[h]Answer options were yes or no; therefore, there are no median and range values.
[i]Alert means no intention to breastfeed.
[j]Alert means feeding baby only formula.
[k]T3: maternity week.

Table 5. Outcomes per patient-reported experience measure domain.

Domain and subdomains	Time point	Value, n (%)	Score, median (range)	Alerts, n (%)	Missing[a], n (%)
Satisfaction with care					
	All	982 (51.07)	3 (1-4)	4 (0.43)	58 (5.91)
	T2[b]	793 (80.75)	3 (1-4)	4 (0.53)	45 (5.67)
	T4[c]	170 (17.31)	4 (2-4)	0 (0)	13 (7.64)
	T5[d]	19 (1.93)	3 (2-4)	0 (0)	0 (0)
Health care responsiveness and shared decision-making					
	All	982 (51.07)	16 (2-16)	101 (10.67)	35 (3.56)
	T2	793 (80.75)	16 (2-16)	82 (10.72)	28 (3.53)
	T4	170 (17.31)	16 (2-16)	17 (10.43)	7 (4.12)
	T5	19 (1.93)	14 (4-16)	2 (10.53)	0 (0)
Birth experience	All (T4)	170 (8.84)	30 (8-40)	37 (23.27)	11 (6.47)
Pain relief					
Information ante partum	All (T2)	793 (41.24)	1 (0-2)	310 (41.33)	43 (5.42)
Experience at birth	All (T4)	170 (8.84)	3 (1-4)	4 (2.65)	19 (11.18)
Partner role					
During pregnancy	All (T2)	793 (41.24)	3 (0-5)	56 (7.35)	31 (3.91)
At birth	All (T4)	170 (8.84)	4 (0-5)	1 (0.66)	18 (10.59)
Continuity of care					
	All	963 (50.08)	11 (4-12)	55 (6.08)	58 (6.02)
	T2	793 (82.35)	11 (4-12)	49 (6.54)	44 (5.55)
	T4	170 (17.65)	11 (4-12)	6 (3.85)	14 (8.24)

[a]Completely missing.
[b]T2: early third trimester.
[c]T4: 6 weeks post partum.
[d]T5: 6 months post partum

PROM per Domain

Social Support

Of the 1092 women who were asked the social support question, administered at the first time point in pregnancy that each site had implemented, 44 (4.03%) scored an alert, meaning that they had 1 or no person near them to count on in time of difficulty. A comparison of T1 and T2 showed a slightly higher alert rate at T2 (17/25, 6.8%) than at T1 (26/815, 3.19%).

Quality of Life

The quality-of-health domain, assessed with the Patient-Reported Outcomes Measurement Information System–Global Health Short Form, had few alerts at all time points. The alerts were based on the sum score; no alerts came from a high pain score. In additional analysis, calculation of subscores for mental and physical health showed no variation across time points.

Mental Health

In 3.52% (61/1731) of the women completing the 2-item depression screening (Patient Health Questionnaire-2 [PHQ-2]) an alert was scored, without variations over time. Women with an alert on the PHQ-2 filled out the full depression questionnaire (ie, Edinburgh Postnatal Depression Scale-10 [EPDS-10]). As 1 region dismissed the PHQ-2 screening questions, 29 women filled out the EPDS-10 directly. The EPDS-10 exceeded the clinical threshold in 52% (47/90) of the women, meaning that 2.67% (47/1760) of the women in the whole population screened positive for depression. The numbers with regard to the EPDS-10 results were too small to allow for interpreting variations over time.

Incontinence and Dyspareunia

The screening question for urine and flatus incontinence was positive in 1 of 4 women. This proportion was lower at T1 than at the other time points. Screening for stool incontinence was positive in 0.86% (15/1741) of the cases, mostly at T4 (6/162, 3.7%). The full questionnaires in case of a positive incontinence screening resulted in alert rates of 39.4% (185/469) on urine incontinence (International Consultation on Incontinence Questionnaire, Short Form) and 97.96% (385/393) on flatus or stool incontinence or both (Wexner scale). Women who screened positive for flatus incontinence but not to stool incontinence scored lower on the Wexner scale (median 3; range 0-11) than women who screened positive for stool incontinence with or without flatus incontinence (median 6; range 1-17). In 24.7% (229/929) of the women, an alert was scored on dyspareunia, with a lower alert rate at T1 than at the other time points.

Breastfeeding

During pregnancy, 77.6% (596/768) of the women intended to breastfeed their baby. After giving birth, 64% (80/125) of the women indicated that they would breastfeed their baby (fully or combined with formula) in the first week post partum, which decreased over time: 59% (89/150) at 6 weeks and 47% (9/19) at 6 months post partum. Of the 761 women who were breastfeeding (T3 or T4) or intended to (T2), 175 (23%) scored an alert on the screening question for confidence in breastfeeding. This alert rate was higher during pregnancy than during the postpartum period. After a positive screening question, the full breastfeeding self-efficacy questionnaire (ie, Breastfeeding Self-Efficacy Scale-10) gave an alert in 72.9% (124/170) of the cases.

Role Transition

The mother-child bonding questionnaire (Mother-to-Infant Bonding Scale) had a median score of 2 (range 0-11) and 44.9% (122/272) alert values. No difference was seen over time. The single question about confidence in the role as mother scored almost no alerts, and the median score was equal to the maximum score.

PREM per Domain

Individual Insight Into PREMs

Before answering PREM questionnaires at T2 (early third trimester), the women could choose whether to give their care professional direct insight into their answers because the answers could affect the dependent relationship with their care professional. The answer to this question was not reported by all participating sites. We received data of 175 women, of whom 26 (14.9%) did not agree to share the answers of their PREM questionnaire directly with their caregiver.

Satisfaction With Care

This single-question domain, filled out by 924 women, scored almost no alerts, and the median score was 3 out of 4 (range 1-4).

Health Care Responsiveness and Shared Decision-making

Total scores were high, with a median of 16 (range 2-16) without variation over time. Still, the alert rate for this domain was 10.7% (101/947), based on a negative answer to one or more questions. Of the 101 women scoring an alert, 59 (58.4%) answered in the negative to just 1 of 8 questions. The alerts per question provided insight into direction for improvement, such as information provision about care decisions.

Birth Experience

Assessed with the 10-item Birth Satisfaction Scale, Revised, at T4, this domain gave an alert in 23.3% (37/159) of the women and had a median total score of 30 (range 8-40). The Birth Satisfaction Scale, Revised, subscales scored a median of 11 (range 2-16) for stress, 14 (range 4-16) for quality of care, and 5 (range 0-8) for women's attributes. Comparing women with and without an alert on the sum score, the subscales stress and women's attributes decreased by 50%, whereas the subscale quality of care decreased by 21%.

Pain Relief

During pregnancy, at T2, 41.3% (310/750) of the women indicated that the options for pain relief had not been discussed with their care professional yet. Post partum, most women were satisfied with the options for pain relief that were offered during childbirth.

Partner Role

Women were asked whether care professionals had engaged their partner enough in their care. This was insufficient for 7.4% (56/762) of the women during pregnancy and for 0.7% (1/152) during labor.

Continuity of Care

In total, 6.1% (55/905) of the women answered in the negative to one or more questions about continuity of care, with a median score of 11 (range 4-12). This domain had a slightly higher alert rate in pregnancy than during the postpartum period. In 96% (53/55) of the alerts, the women scored only 1 of the 3 questions negatively. Most alerts resulted from a negative answer to the question about knowing who their principal care provider was. In 23.5% (213/905) of the cases, the women had received perinatal care from just 1 care professional. Excluding these, the overall alert rate was 7.9% (55/692) and the median score 10 (range 4-12).

Adherence to the Questionnaires

Overall, 84% (1615/1923) of the questionnaires were filled out completely. Per domain, the percentage of completely missing answers ranged between 0% and 13%, as presented in Tables 4 and 5. Certain domains were skipped more often, such as the EPDS-10 (depression) and the Patient-Reported Outcomes Measurement Information System–Sexual Function and Satisfaction (PROMIS-SFFAC102; pain with intercourse). Missing rates per question are listed in Multimedia Appendix 3 and ranged from 0% to 16%. Evaluated per question, no remarkable missing patterns were found that could not be explained by site-specific adaptations to the questions. In Multimedia Appendix 4, missing patterns per domain are visualized. In additional sensitivity analysis of domains with multiple questions, sum scores and alert rates did not significantly change after ruling out the questionnaires with >25% missing items. Here, we chose to report the complete case analysis, best reflecting clinical use, because these results were not ruled out from individual reports to care professionals.

Discussion

Findings and Recommendations

This study reports the results of an innovation in perinatal care in the Netherlands: implementation of ICHOM's PROM and PREM domains for pregnancy and childbirth to guide individual patient care in 7 OCNs. The large cohort resulting from this project showed good adherence to the questionnaires. In several domains, such as incontinence and breastfeeding, the high alert rates revealed opportunities to improve and personalize perinatal care for individual women on outcomes that matter to them. In addition, our results indicate that some measurement instruments and their timing as proposed by ICHOM are less suitable for clinical use. On the basis of these findings, we present several recommendations regarding the methods and timelines of PROM and PREM assessment in clinical practice.

Overall, adherence to the questionnaires was good, similar to PROM adherence when used for routine oncologic care [7]. High missing rates per instrument could be explained by technical issues, site-specific adaptation to the questionnaires, or questions addressing a relatively taboo subject, such as those included in the EPDS-10 and PROMIS-SFFAC102 (depression and pain with intercourse, respectively). In preimplementation tests, the PROMIS-SFFAC102 question also seemed difficult to understand despite language adjustments. Adapting the answer options might help, or an alternative instrument should be selected. Although they may be imperfect, the questions on these taboo subjects were answered by most women. Especially, these taboo subjects create more awareness at both patient and care professional levels, thereby increasing the likelihood of problems being recognized and addressed in clinic.

Median moments of completion corresponded well with the timeline of data collection as proposed by ICHOM. In contrast to the provider expectations described by Chen et al [27], the questionnaire administered shortly after childbirth (T3) resulted in a large group of respondents in this study who completed them mostly within 2 weeks post partum. At this point, there is an excellent opportunity to improve breastfeeding outcomes and mother-child bonding. As final maternal checkup with an obstetric care professional is at 6 weeks post partum in the Netherlands, the questionnaire at 6 months post partum (T5) is practically difficult to arrange for care providers. As a result, most OCNs chose to skip T5 to enhance feasibility; thus, few questionnaires were collected. Although practically challenging, patient views on this timing should be considered because this moment previously has been shown to be valuable to reflect on long-term recovery after pregnancy and childbirth [17,28].

Our findings in the mental health domain indicate that the first instrument of the 2-step screening (PHQ-2) is missing an unacceptable proportion of women at risk for depression, in line with the findings of Slavin et al [21]. The prevalence of perinatal depression has been reported at a rate of 7% to 20% during pregnancy and up to 22% in the first year post partum [29]. In our cohort, the prevalence of depressive symptoms was only 2.7% over the whole period of pregnancy and childbirth up until 6 months post partum. As the main purpose in clinical care is to identify women at high risk for depression, we strongly recommend removing the PHQ-2 and screening all women for depressive complaints with the EPDS-10, despite an increased response burden. The EPDS-10 has been thoroughly validated and has been shown to be acceptable to women in pregnancy and post partum [30,31]. Furthermore, 2 PREM domains showed striking results. Women answered almost always in the positive to the PREM *satisfaction with results of care*, despite multiple PROM alerts suggesting that their results were not as positive. This might be explained by women expecting incontinence to be a *normal* result of pregnancy and childbirth. Either way, this single question did not differentiate between women who were satisfied and those who were unsatisfied with their care and does not add value to shared decision-making or quality improvement. The PREM on information provision about pain relief options gave unexpected high alerts: 41.3% (310/750) of the care professionals had not discussed this yet with their patient. This might indicate that the timing of the assessment does not fit clinical practice because the T2 questionnaire was completed at 28 weeks of pregnancy on average and regular pathways plan to discuss pain relief later. Overall, each domain

in need of adjustment based on our results is listed in Textbox 2, along with proposed adaptations to enhance their use in clinical practice.

In several domains, high alert rates revealed opportunities to adapt care accordingly and improve individual outcomes. For example, a high prevalence of incontinence and pain with intercourse was found over the course of pregnancy, as expected from previous research on these topics [32]. Breastfeeding success rates were low, which corresponds to provider-reported breastfeeding numbers in the Netherlands from 2018 [33]. Strikingly, many alerts were scored on breastfeeding confidence and self-efficacy during pregnancy. This provides important opportunities for all perinatal care professionals involved to improve breastfeeding outcomes. At the same time, threshold values for alerts on several instruments must be evaluated for clinical use to determine whether women scoring an alert want help and whether clinicians have the instruments to provide this help. For example, the threshold for the Mother-to-Infant Bonding Scale was set quite low based on the literature [34,35], resulting in many alerts on mother-child bonding. At this moment, it is unknown whether women want their care professional to address these alerts, and clinical guidelines on when and how to act are lacking [36]. However, in perinatal care too, structural PROM monitoring did create openings for dialogue between patients and care professionals to personalize and improve care on these themes [2].

Regarding experience domains, 85.1% (149/175) of the women in this study agreed to making their individual answers to PREMs visible to their care professionals, but the remaining 14.9% (26/175) disagreed. These numbers both affirm the acceptability of individual PREM use and underline the importance of providing women an opportunity to choose, considering their dependent relationship with care professionals. In general, evaluating results of all women, the sum scores of the PREM instruments often did not differentiate very much, but separate answers gave valuable information about directions for improvement. For example, most alerts in the domains continuity and health care responsiveness resulted from negative answers to specific items: about knowing their principal care professional and information provision, respectively. In birth experience, the PREM with the highest alert rate, the subscales most affected in women with an alert on the sum score were stress and women's attributes. Until now, the literature on individual PREM use to guide clinical practice has been scarce because anonymous use is mostly advocated, for quality improvement only [17,37].

Textbox 2. Proposed adaptations to pregnancy and childbirth set content.

Mental health

- Remove Patient Health Questionnaire-2 and use only the Edinburgh Postnatal Depression Scale-10 to screen depressive symptoms because current 2-step screening rules out too many women at risk for perinatal depression.

Incontinence

- Use the first question of the International Consultation on Incontinence Questionnaire, Short Form, and first 3 questions of the Wexner scale as screening questions because they ask the same questions as the current screening questions. The current screening questions create an unnecessary response burden and have led to inconsequential answers.

Pain with intercourse

- Adjust the answer options or replace the instrument considering its relatively high missing rate and signs that the question is hard to understand.

Role as mother

- Replace with another instrument because this single question does not differentiate between women who were confident and those who were insecure in their role as mother. As patients proposed this subject originally, it should be maintained in the pregnancy and childbirth set [16].

Satisfaction with care

- Remove or replace with another instrument because this question does not differentiate between women who were satisfied and those who were unsatisfied with their care or provide insight into the direction for improvements.

Pain relief

- Measurement at T2 (early third trimester) is often too early because most perinatal care professionals discuss pain relief options later in the care path. We recommend involving patients to determine the optimal timing in pregnancy to discuss options for pain relief during childbirth.

Social support

- Ask it at each time point because women's social networks can change throughout pregnancy and post partum. This domain was originally designed as a casemix factor but is used in clinical practice also as an outcome to act upon.

Before asking questions about patient experiences

- Ask the woman whether her answers to the patient-reported experience measure questions may be made visible to her care professional individually because women are in a dependent relationship with their care professionals.

Strengths and Limitations

To our knowledge, this project was one of the first experiences with incorporating the complete PCB set into clinical practice to guide individual perinatal care. Although it was challenging, each participating site collaborated with a multidisciplinary transmural team of care professionals (part of an OCN) for implementation to ensure continuity of care over the whole cycle of care in a patient-centered approach. For this study, we have performed thorough additional analyses such as sensitivity analysis and appraisal of the use of screening questions, leading to practice implications for several domains. The sample size was large, and our results reflect the true clinical use of all patient-reported domains in the PCB set in various settings across the Netherlands. Nevertheless, because of this practical and local approach, nonresponders were not registered; therefore, we cannot report any response rates. In addition, variation over time in our results should be interpreted with caution because of different numbers of results per time point—especially, the numbers at 6 months post partum were too small to enable drawing any conclusions. Another limitation was the absence of questionnaire translations, restricting the participants to Dutch-speaking women only. Moreover, because no resources were available to support completion of the questionnaires, women with low (digital) health literacy are likely to be underrepresented, although women with language barriers or low health literacy probably have higher prevalence of pregnancy-related issues and thus greater opportunities to improve their outcomes [38]. This reveals an important concern regarding the transformation to value-based care: it could worsen existing health inequities even further. Therefore, efforts should be made to standardize the questionnaires to facilitate translation into multiple languages. Furthermore, when implementing PROMs and PREMs as part of value-based care, all stakeholders involved should be well informed about their purpose and supported with multiple solutions to embed the PCB set structurally in clinic; for example, through group consultations [39].

Implications for Practice

On the basis of the first efforts to incorporate the PCB set into clinical practice, we have proposed several adaptations to its content and structure to better fit routine perinatal care (Textbox 2). At the same time, international governance of the PCB set is essential to maintain comparability for care improvement purposes. In addition, although we tested their clinical usefulness, further validation is needed of all the measurement instruments and their clinical thresholds during pregnancy and post partum, which has been started successfully in another cohort [20-22]. Although the numbers per region could not be compared because of differences in pilot setup (eg, patient group selection), data capture was more feasible when PROMs could be embedded in their own EHR. When used in performance management, PROM and PREM results would preferably be merged with clinical outcomes, ideally through the EHR. Although beyond our main scope, merging patient-reported data with clinical outcomes from EHRs was explored in this project. In concordance with previous findings [40], this seemed very challenging, depending on the software systems available. This study focused on the content of the PCB set; future work should investigate other factors influencing implementation in the patient, care professional, and organization contexts [41].

Conclusions

This study shows that the PCB set is a useful tool to capture and discuss patient-reported outcomes and experiences that need attention during pregnancy, childbirth, and post partum. These are promising findings in the journey toward patient-centered, personalized, and value-based perinatal care. In the future, merging patient-reported data with clinical outcomes and casemix factors would be even more valuable to improve quality of health care both at an individual level and an aggregated level.

Acknowledgments

The authors thank all patients who gave consent to use their questionnaires for research. The authors acknowledge the local project teams and care professionals for their implementation efforts. This work was supported by Zorginstituut Nederland (2018026697). Zorginstituut Nederland was not involved in the study design, data collection, analysis and interpretation of data, writing the report, or decision to submit the article for publication.

The contributors associated with BUZZ Project Team are as follows: Jolanda H. Vermolen, Maternity care organization De Waarden, Schoonhoven, the Netherlands; Simone A. Vankan-Buitelaar, Department of Obstetrics and Gynecology, Division of Obstetrics and Fetal Medicine, Erasmus MC Sophia, Rotterdam, the Netherlands; Monique Klerkx, Integrated obstetric care organization Annature Geboortezorg, Breda, the Netherlands; Pieter-Kees de Groot, Department of Obstetrics, Spaarne Gasthuis, Haarlem, the Netherlands; Nikkie Koper, Midwifery practice Duo, Zwanenburg, the Netherlands; Hilmar H. Bijma, Department of Obstetrics and Gynecology, Division of Obstetrics and Fetal Medicine, Erasmus MC Sophia, Rotterdam, Netherlands; Bas van Rijn, Department of Obstetrics and Gynecology, Division of Obstetrics and Fetal Medicine, Erasmus MC Sophia, Rotterdam, the Netherlands; Elise Neppelenbroek, Midwifery practice Verloskundigen Bakerraad, Zwolle, the Netherlands; Sebastiaan W.A. Nij Bijvank, Department of Obstetrics, Isala Clinics, Zwolle, the Netherlands; Marieke Veenhof, Department of Obstetrics, Leiden University Medical Center, Leiden, the Netherlands; Margot Lutke Holzik, Department of Obstetrics, Leiden University Medical Center, Leiden, the Netherlands; Jelle H. Baalman, Department of Obstetrics, Medisch Spectrum Twente, Enschede, the Netherlands; Josien Vrielink-Braakman, Midwifery practice Verwacht, Vriezenveen, the Netherlands.

Authors' Contributions

ALD, MLdR, LTL, AF, HEES, and MNB designed the study. The BUZZ team performed the data collection in clinic. ALD and MLdR analyzed the data under the supervision of MNB. All authors interpreted the data. ALD wrote the first version of the manuscript under the supervision of MLdR and MNB. All authors revised all versions of the manuscript and approved the final version.

Conflicts of Interest

AF and MLdR were part of the International Consortium for Health Outcomes Measurement Pregnancy and Childbirth Outcome Set Working Group.

Multimedia Appendix 1
Overview of the pregnancy and childbirth outcome set: structure, scoring, and alert values.
[XLSX File (Microsoft Excel File), 32 KB - jmir_v24i7e37725_app1.xlsx]

Multimedia Appendix 2
Outcomes per domain and time point.
[PDF File (Adobe PDF File), 71 KB - jmir_v24i7e37725_app2.pdf]

Multimedia Appendix 3
Missing rates per question.
[XLSX File (Microsoft Excel File), 12 KB - jmir_v24i7e37725_app3.xlsx]

Multimedia Appendix 4
Missing patterns per domain.
[PDF File (Adobe PDF File), 30 KB - jmir_v24i7e37725_app4.pdf]

References

1. Steinmann G, van de Bovenkamp H, de Bont A, Delnoij D. Redefining value: a discourse analysis on value-based health care. BMC Health Serv Res 2020 Sep 14;20(1):862 [FREE Full text] [doi: 10.1186/s12913-020-05614-7] [Medline: 32928203]
2. Gray M, Gray J, Howick J. Personalised healthcare and population healthcare. J R Soc Med 2018 Feb 18;111(2):51-56 [FREE Full text] [doi: 10.1177/0141076817732523] [Medline: 28920755]
3. Austin E, LeRouge C, Hartzler AL, Segal C, Lavallee DC. Capturing the patient voice: implementing patient-reported outcomes across the health system. Qual Life Res 2020 Feb 12;29(2):347-355 [FREE Full text] [doi: 10.1007/s11136-019-02320-8] [Medline: 31606809]
4. Damman OC, Jani A, de Jong BA, Becker A, Metz MJ, de Bruijne MC, et al. The use of PROMs and shared decision-making in medical encounters with patients: an opportunity to deliver value-based health care to patients. J Eval Clin Pract 2020 Apr;26(2):524-540 [FREE Full text] [doi: 10.1111/jep.13321] [Medline: 31840346]
5. Black N. Patient reported outcome measures could help transform healthcare. BMJ 2013 Jan 28;346:f167. [doi: 10.1136/bmj.f167] [Medline: 23358487]
6. Patient-reported outcomes. National Quality Forum. URL: https://www.qualityforum.org/Patient-Reported_Outcomes.aspx [accessed 2021-12-23]
7. van Egdom LS, Oemrawsingh A, Verweij LM, Lingsma HF, Koppert LB, Verhoef C, et al. Implementing patient-reported outcome measures in clinical breast cancer care: a systematic review. Value Health 2019 Oct;22(10):1197-1226 [FREE Full text] [doi: 10.1016/j.jval.2019.04.1927] [Medline: 31563263]
8. Basch E, Barbera L, Kerrigan CL, Velikova G. Implementation of patient-reported outcomes in routine medical care. Am Soc Clin Oncol Educ Book 2018 May 23;38:122-134 [FREE Full text] [doi: 10.1200/EDBK_200383] [Medline: 30231381]
9. McAllister M, Dearing A. Patient reported outcomes and patient empowerment in clinical genetics services. Clin Genet 2015 Aug 11;88(2):114-121. [doi: 10.1111/cge.12520] [Medline: 25307491]
10. Basch E. Patient-reported outcomes - harnessing patients' voices to improve clinical care. N Engl J Med 2017 Jan 12;376(2):105-108. [doi: 10.1056/NEJMp1611252] [Medline: 28076708]
11. Scheerhagen M, van Stel HF, Tholhuijsen D, Birnie E, Franx A, Bonsel G. Applicability of the ReproQ client experiences questionnaire for quality improvement in maternity care. PeerJ 2016;4(7):e2092 [FREE Full text] [doi: 10.7717/peerj.2092] [Medline: 27478690]
12. 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]

13. Foster A, Croot L, Brazier J, Harris J, O'Cathain A. The facilitators and barriers to implementing patient reported outcome measures in organisations delivering health related services: a systematic review of reviews. J Patient Rep Outcomes 2018 Dec 3;2(1):46 [FREE Full text] [doi: 10.1186/s41687-018-0072-3] [Medline: 30363333]
14. Nordan L, Blanchfield L, Niazi S, Sattar J, Coakes CE, Uitti R, et al. Implementing electronic patient-reported outcomes measurements: challenges and success factors. BMJ Qual Saf 2018 Oct 18;27(10):852-856. [doi: 10.1136/bmjqs-2018-008426] [Medline: 30021802]
15. Data collection results. In: Workplace Environmental Design in Architecture for Public Health. Cham: Springer; 2017.
16. Nijagal MA, Wissig S, Stowell C, Olson E, Amer-Wahlin I, Bonsel G, et al. Standardized outcome measures for pregnancy and childbirth, an ICHOM proposal. BMC Health Serv Res 2018 Dec 11;18(1):953 [FREE Full text] [doi: 10.1186/s12913-018-3732-3] [Medline: 30537958]
17. Laureij LT, Been JV, Lugtenberg M, Ernst-Smelt HE, Franx A, Hazelzet JA, PCB outcome set study group Collaborative authors, et al. Exploring the applicability of the pregnancy and childbirth outcome set: a mixed methods study. Patient Educ Couns 2020 Mar;103(3):642-651 [FREE Full text] [doi: 10.1016/j.pec.2019.09.022] [Medline: 31607633]
18. Al-Shammari I, Roa L, Yorlets RR, Akerman C, Dekker A, Kelley T, et al. Implementation of an international standardized set of outcome indicators in pregnancy and childbirth in Kenya: utilizing mobile technology to collect patient-reported outcomes. PLoS One 2019 Oct 16;14(10):e0222978 [FREE Full text] [doi: 10.1371/journal.pone.0222978] [Medline: 31618249]
19. Depla AL, Ernst-Smelt HE, Poels M, Crombag NM, Franx A, Bekker MN. A feasibility study of implementing a patient-centered outcome set for pregnancy and childbirth. Health Sci Rep 2020 Sep 26;3(3):e168 [FREE Full text] [doi: 10.1002/hsr2.168] [Medline: 32607452]
20. Slavin V, Creedy DK, Gamble J. Single item measure of social supports: evaluation of construct validity during pregnancy. J Affect Disord 2020 Jul 01;272:91-97. [doi: 10.1016/j.jad.2020.03.109] [Medline: 32379626]
21. Slavin V, Creedy DK, Gamble J. Comparison of screening accuracy of the Patient Health Questionnaire-2 using two case-identification methods during pregnancy and postpartum. BMC Pregnancy Childbirth 2020 Apr 14;20(1):211 [FREE Full text] [doi: 10.1186/s12884-020-02891-2] [Medline: 32290813]
22. Slavin V, Gamble J, Creedy DK, Fenwick J, Pallant J. Measuring physical and mental health during pregnancy and postpartum in an Australian childbearing population - validation of the PROMIS Global Short Form. BMC Pregnancy Childbirth 2019 Oct 22;19(1):370 [FREE Full text] [doi: 10.1186/s12884-019-2546-6] [Medline: 31640626]
23. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. Int J Surg 2014 Dec;12(12):1495-1499 [FREE Full text] [doi: 10.1016/j.ijsu.2014.07.013] [Medline: 25046131]
24. Outcome based healthcare 2018-2022. Government of Netherlands. URL: https://www.government.nl/documents/reports/2018/07/02/outcome-based-healthcare-2018-2022 [accessed 2022-01-10]
25. Perinatale zorg in Nederland anno 2020. Perined. URL: https://assets.perined.nl/docs/3d6a2b46-aa8a-417e-a55e-de0184fe2078.pdf [accessed 2022-01-10]
26. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2020.
27. Chen A. Introducing patient reported measures (PRMs) into Finnish public maternity service: a study on professionals' views. BMC Health Serv Res 2021 (forthcoming). [doi: 10.21203/rs.3.rs-129756/v1]
28. Depla AL, Crombag NM, Franx A, Bekker MN. Implementation of a standard outcome set in perinatal care: a qualitative analysis of barriers and facilitators from all stakeholder perspectives. BMC Health Serv Res 2021 Feb 02;21(1):113 [FREE Full text] [doi: 10.1186/s12913-021-06121-z] [Medline: 33530989]
29. Littlewood E, Ali S, Dyson L, Keding A, Ansell P, Bailey D, et al. Identifying perinatal depression with case-finding instruments: a mixed-methods study (BaBY PaNDA – Born and Bred in Yorkshire PeriNatal Depression Diagnostic Accuracy). Health Serv Deliv Res 2018 Feb;6(6):1-210. [doi: 10.3310/hsdr06060] [Medline: 29481018]
30. Hewitt C, Gilbody S, Brealey S, Paulden M, Palmer S, Mann R, et al. Methods to identify postnatal depression in primary care: an integrated evidence synthesis and value of information analysis. Health Technol Assess 2009 Jul;13(36):1-145, 147 [FREE Full text] [doi: 10.3310/hta13360] [Medline: 19624978]
31. Bergink V, Kooistra L, Lambregtse-van den Berg MP, Wijnen H, Bunevicius R, van Baar A, et al. Validation of the Edinburgh Depression Scale during pregnancy. J Psychosom Res 2011 Apr;70(4):385-389. [doi: 10.1016/j.jpsychores.2010.07.008] [Medline: 21414460]
32. Brincat C, Crosby E, McLeod A, Fenner DE. Experiences during the first four years of a postpartum perineal clinic in the USA. Int J Gynaecol Obstet 2015 Jan 06;128(1):68-71. [doi: 10.1016/j.ijgo.2014.07.026] [Medline: 25262441]
33. Rapport Peiling Melkvoeding 2018. NCJ. URL: https://assets.ncj.nl/docs/4353efb6-135e-4ef4-8afb-a1b6f772995c.pdf [accessed 2022-01-10]
34. Bienfait M, Maury M, Haquet A, Faillie J, Franc N, Combes C, et al. Pertinence of the self-report mother-to-infant bonding scale in the neonatal unit of a maternity ward. Early Hum Dev 2011 Apr;87(4):281-287. [doi: 10.1016/j.earlhumdev.2011.01.031] [Medline: 21316876]

35. Yoshida K, Yamashita H, Conroy S, Marks M, Kumar C. A Japanese version of Mother-to-Infant Bonding Scale: factor structure, longitudinal changes and links with maternal mood during the early postnatal period in Japanese mothers. Arch Womens Ment Health 2012 Oct 26;15(5):343-352 [FREE Full text] [doi: 10.1007/s00737-012-0291-1] [Medline: 22733162]
36. Woodhouse SS, Scott JR, Hepworth AD, Cassidy J. Secure base provision: a new approach to examining links between maternal caregiving and infant attachment. Child Dev 2020 Jan 11;91(1):e249-e265. [doi: 10.1111/cdev.13224] [Medline: 30740649]
37. Lagha E, Noble A, Smith A, Denvir M, Leslie S. Patient Reported Experience Measures (PREMs) in chronic heart failure. J R Coll Physicians Edinb 2012 Dec 14;42(4):301-305. [doi: 10.4997/jrcpe.2012.404]
38. Vos AA, Denktaş S, Borsboom GJ, Bonsel GJ, Steegers EA. Differences in perinatal morbidity and mortality on the neighbourhood level in Dutch municipalities: a population based cohort study. BMC Pregnancy Childbirth 2015 Sep 02;15(1):201 [FREE Full text] [doi: 10.1186/s12884-015-0628-7] [Medline: 26330115]
39. Carter EB, EleVATE Women Collaborative, Mazzoni SE. A paradigm shift to address racial inequities in perinatal healthcare. Am J Obstet Gynecol 2021 Apr;224(4):359-361. [doi: 10.1016/j.ajog.2020.11.040] [Medline: 33306974]
40. Aerts H, Kalra D, Sáez C, Ramírez-Anguita JM, Mayer M, Garcia-Gomez JM, et al. Quality of hospital electronic health record (EHR) data based on the international consortium for health outcomes measurement (ICHOM) in heart failure: pilot data quality assessment study. JMIR Med Inform 2021 Aug 04;9(8):e27842 [FREE Full text] [doi: 10.2196/27842] [Medline: 34346902]
41. Roberts NA, Alexander K, Wyld D, Janda M. Factors in randomized controlled trials reported to impact the implementation of patient-reported outcome measures into routine care: protocol for a systematic review. JMIR Res Protoc 2019 Nov 26;8(11):e14579 [FREE Full text] [doi: 10.2196/14579] [Medline: 31769763]

Abbreviations

BUZZ: Dutch abbreviation of Discuss Outcomes of Pregnancy with the Pregnant Woman
EHR: electronic health record
EPDS-10: Edinburgh Postnatal Depression Scale-10
ICHOM: International Consortium for Health Outcomes Measurement
OCN: obstetric care network
PCB set: set of patient-centered outcome measures for pregnancy and childbirth
PHQ-2: Patient Health Questionnaire-2
PREM: patient-reported experience measure
PROM: patient-reported outcome measure
PROMIS-SFFAC102: Patient-Reported Outcomes Measurement Information System–Sexual Function and Satisfaction

Edited by T Leung; submitted 04.03.22; peer-reviewed by A Nicolas-Boluda, W Leung; comments to author 23.04.22; revised version received 04.05.22; accepted 31.05.22; published 05.07.22.

Please cite as:
Depla AL, Lamain-de Ruiter M, Laureij LT, Ernst-Smelt HE, Hazelzet JA, Franx A, Bekker MN, BUZZ Project Team
Patient-Reported Outcome and Experience Measures in Perinatal Care to Guide Clinical Practice: Prospective Observational Study
J Med Internet Res 2022;24(7):e37725
URL: https://www.jmir.org/2022/7/e37725
doi:10.2196/37725
PMID:35787519

Original Paper

Drug Recommendation System for Diabetes Using a Collaborative Filtering and Clustering Approach: Development and Performance Evaluation

Luis Fernando Granda Morales[1], MSc; Priscila Valdiviezo-Diaz[1], PhD; Ruth Reátegui[1], PhD; Luis Barba-Guaman[1], PhD

Departamento de Ciencias de la Computación y Electrónica, Universidad Técnica Particular de Loja, Loja, Ecuador

Corresponding Author:
Priscila Valdiviezo-Diaz, PhD
Departamento de Ciencias de la Computación y Electrónica
Universidad Técnica Particular de Loja
San Cayetano Alto
Loja, 1101608
Ecuador
Phone: 593 7 3701444 ext 2325
Email: pmvaldiviezo@utpl.edu.ec

Abstract

Background: Diabetes is a public health problem worldwide. Although diabetes is a chronic and incurable disease, measures and treatments can be taken to control it and keep the patient stable. Diabetes has been the subject of extensive research, ranging from disease prevention to the use of technologies for its diagnosis and control. Health institutions obtain information required for the diagnosis of diabetes through various tests, and appropriate treatment is provided according to the diagnosis. These institutions have databases with large volumes of information that can be analyzed and used in different applications such as pattern discovery and outcome prediction, which can help health personnel in making decisions about treatments or determining the appropriate prescriptions for diabetes management.

Objective: The aim of this study was to develop a drug recommendation system for patients with diabetes based on collaborative filtering and clustering techniques as a complement to the treatments given by the treating doctor.

Methods: The data set used contains information from patients with diabetes available in the University of California Irvine Machine Learning Repository. Data mining techniques were applied for processing and analysis of the data set. Unsupervised learning techniques were used for dimensionality reduction and patient clustering. Drug predictions were obtained with a user-based collaborative filtering approach, which enabled creating a patient profile that can be compared with the profiles of other patients with similar characteristics. Finally, recommendations were made considering the identified patient groups. The performance of the system was evaluated using metrics to assess the quality of the groups and the quality of the predictions and recommendations.

Results: Principal component analysis to reduce the dimensionality of the data showed that eight components best explained the variability of the data. We identified six groups of patients using the clustering algorithm, which were evenly distributed. These groups were identified based on the available information of patients with diabetes, and then the variation between groups was examined to predict a suitable medication for a target patient. The recommender system achieved good results in the quality of predictions with a mean squared error metric of 0.51 and accuracy in the quality of recommendations of 0.61, which is acceptable.

Conclusions: This work presents a recommendation system that suggests medications according to drug information and the characteristics of patients with diabetes. Some aspects related to this disease were analyzed based on the data set used from patients with diabetes. The experimental results with clustering and prediction techniques were found to be acceptable for the recommendation process. This system can provide a novel perspective for health institutions that require technologies to support health care personnel in the management of diabetes treatment and control.

(J Med Internet Res 2022;24(7):e37233) doi:10.2196/37233

KEYWORDS

clustering; collaborative filtering; diabetes; recommender system; recommend; drug; chronic disease; patient information; data mining; machine learning

Introduction

Background

Owing to the large amount of information available in health institution databases, including medical treatments, diagnostic tests, clinical histories, and drug characteristics, there is a need to implement recommender systems (RSs) that support medical staff in activities related to health control and management. The main concept of an RS is to suggest items that are particularly suitable for the user based on their profile or historical preferences. In the context of health, these items can be drugs, medical treatments, health videos, and patients sharing the same disease. An RS employs data sources to learn about user preferences through machine-learning algorithms and information-filtering techniques such as content-based, collaborative filtering–based, demographic, and hybrid approaches [1].

Currently, more than 80% of internet users seek health information through various platforms, including social networks, a figure that, according to De Choudhur et al [2], continues to grow. Through the internet, users can identify patients with the same disease, the possible causes of their ailments, find procedures to alleviate a particular disease, learn new healthy habits, and find general health information [3-5].

In the case of chronic diseases such as diabetes, the prescription of multiple drugs is common; thus, RSs can support the intervention of the treating doctor in determining which drugs to prescribe to a particular patient. Considering the current health status of a patient, their clinical history, medications prescribed in previous periods, specific symptoms, and other characteristics, the system can search for patients with similar parameters in the database and suggest drugs that have been more successful in these cases, which could be recommended to the target patient.

Some research has been performed on RSs in the health area. For example, Zhang et al [6] proposed the iDoctor system to provide users with personalized medical recommendations. This system explores users' emotions and preferences about doctors through their ratings and reviews. Gujar et al [7] proposed data-mining techniques for the prediction of a disease and to make a recommendation for specialists about the predicted disease. Kuanr et al [8] proposed an RS for cervical cancer in which predictive models showed high accuracy. Poornima [9] presented a daily nutrition RS for women taking into consideration physical data, preferences, and personal information to combat diseases such as malnutrition, obesity, and cardiovascular diseases.

Other research has focused primarily on recommending doctors and hospitals that are best suited to a specific patient profile [10], medication recommendations [11], treatment recommendations for patients over time [12], videos about health [13], and even customized meal plans [14].

Recently, several studies related to the use of RSs in diabetes have emerged, including some exploratory analyses on the disease, predictions on diet plans to combat obesity and diabetes [14], and physical activity and diet plan RSs to help prevent chronic diseases [15].

Clustering is one of the most widely used machine-learning techniques in the field of health to identify patterns or groups of patients with similar characteristics [16,17]. Although the clustering technique has been the subject of research in the area of RSs, these systems have not yet been widely used in medicine. Moreover, technologies that enable only the analysis or prediction of diagnoses or diseases would not be sufficient to provide personalized care to the patient or to support health personnel in making decisions about which drugs to consider for certain diseases. Therefore, we here propose a drug RS for patients with diabetes based on clustering techniques as a complement to the treatments given by the treating doctor.

Both drug predictions and recommendations were evaluated using traditional metrics to measure the performance of the RS.

Contribution

The aim of this study was to complement previous diabetes-related studies by first analyzing data related to patients with diabetes to obtain important information for the management of this disease, followed by identifying groups of patients who share similar characteristics, which could enable discovering patterns of interest that can support decision-making. Finally, an RS was developed that suggests medications for diabetes according to the patient's historical information and the doses of the medications administered.

Methods

Data Set

The data set used was obtained from the University of California Irvine (UCI) Machine Learning Repository [18], which contains information on patients with diseases associated with diabetes [19]. The original data set includes more than 50 features representing patient outcomes from 130 US hospitals. This data set has more than 100,000 patient records, which refer to 10 years of health care records (from 1999 to 2008).

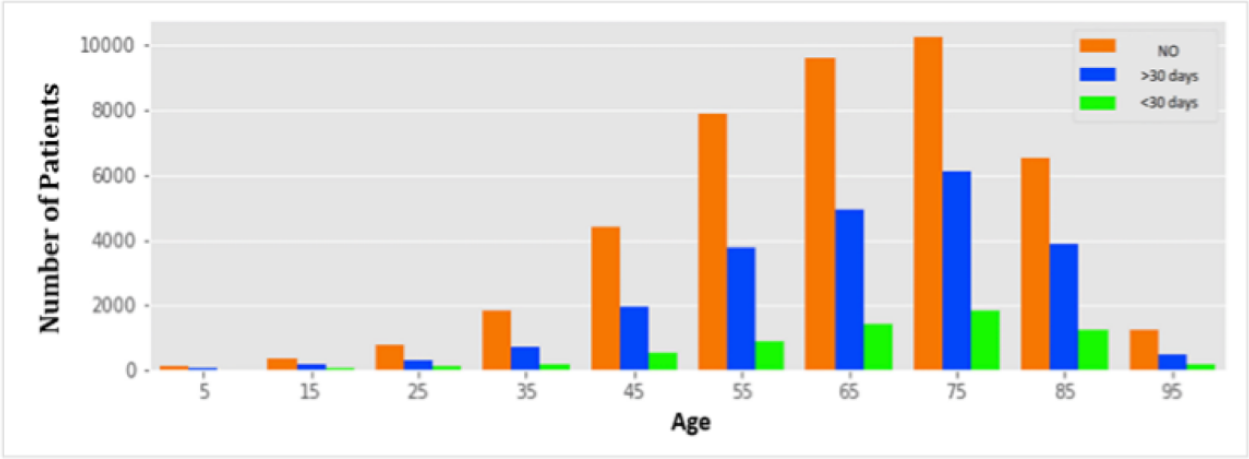
In summary, the information contained in the data set refers to: admissions of patients to the hospital; information on 24 medications administered to patients with diabetes; changes in the patients' medication, and whether the dosage was increased, reduced, kept stable, or not administered; number of medications administered to the patient; time spent by the patient in the hospital, recorded in days; results of tests that were indicated to patients prior to and during their treatment; diagnosis; type of admission; specialty of the treating doctor; and patient data such as age, race, and gender.

Exploratory Data Analysis

Exploratory analysis of the data set is a critical process in research to discover patterns, detect anomalies, test hypotheses, and test assumptions with the help of statistics and graphical representations. It is good practice to first understand the data to obtain as much information as possible.

Toward this end, in the initial analysis of the patient data, we determined that most of the patients belong to the age range of 50 to 80 years, and we classified patients according to readmission status (ie, if the patient presents a case of readmission greater than or less than 30 days, and no readmission). Figure 1 shows the relationship between age and patient readmission, demonstrating very few cases of readmission for younger patients (under 40 years of age). In addition, women had a slightly higher readmission rate than men in cases of readmission longer than 30 days. Readmission showed a similar distribution for patients with and without medication prescribed for diabetes prior to hospital treatment. In addition, we determined that the majority of the patients were of the Caucasian race and did not have a glucose or hemoglobin A1C test.

Figure 1. Distribution of patients according to age and readmission.



Data Preparation

Overview

Data preparation, also known as preprocessing, is a key task in the initial stages to ensure a correct analysis of the information available. Before applying the clustering techniques, the original data set was cleaned by removing all duplicated cases of patients and eliminating records of patients who had not been prescribed any medication. We then performed data transformation and variable selection.

Data Transformation

The data transformation process generally involves converting variables to another type of data and creating new variables. In our case, we converted categorical variables to binary variables, such as gender, maximum glucose serum level, and hemoglobin A1C test result.

New variables were created from the categories of the first diagnosis variable, which was coded according to the first three digits of the International Classification of Diseases-9 system: circulatory, diabetes, digestive, genitourinary, injuries, musculoskeletal, neoplasms, respiratory, and other. The categories of the race variable were created as new variables. Subsequently, the first diagnosis and race variables were removed from the data set. In addition, the age ranges were replaced by the mean of the ranges.

Variable Selection

Noninformative features in the data set were discarded due to a large number of missing values (50,000/100,000, 50.00%) or because some features were not relevant to classifying the data, such as patient identification, or if the feature is unbalanced (n=95,000, >95% of the data had the same value for a feature). In addition, we selected patients who had been prescribed at least two medications. Table 1 lists the discarded parameters (features) and the reasons for discarding them.

As a result, a final data set was obtained with 5177 unique patient records and 42 variables, which were categorized as patient characteristics and medications administered to patients. These final attributes are detailed in Table 2 and Table 3, respectively.

Drugs whose administration represented a very small percentage (500/100,000, 0.50%) or drugs that had not been administered to any patient (as was the case for examide and cytoglipton) were eliminated. Table 3 lists the drugs selected after data processing and the number and proportions of patients using the drug.

For each drug, we classified whether it was administered or if there was a change in the dose. We considered four values for this variable: “up” indicates that the dose was increased during the patient encounter, “down” indicates that the dose was decreased, “steady” indicates that the dose did not change, and “no” indicates that the drug was not administered.

As shown in Table 3, insulin was administered to more than 50% of the patients present in our data set, metformin was

administered to slightly less than 20% of the patients, followed by glipizide and glyburide. The drugs that were administered to fewer patients were glyburide-metformin and nateglinide.

Table 1. Variables discarded from the data set.

Variable	Discard reason
encounter_id[a]	Irrelevant variable for clustering
patient_nbr[b]	Irrelevant variable for clustering
payer_code[c]	Irrelevant variable for clustering
Weight	Data missing for 97.00% (n=97,000) of the 100,000 samples
Medical specialty	Data missing for 53.00% (n=53,000) of the 100,000 samples
Clorpropamida	Only 86 patients use this drug
Acarbosa	Only 308 patients use this drug
Miglitol	Only 38 patients use this drug
Troglitazona	Only 3 patients use this drug
Examide	No patient uses this drug
Citoglipton	No patient uses this drug
Glipizide_metformin	Only 13 patients use this drug
Glimepirida_pioglitazona	Only 1 patient uses this drug
Metformin_rosiglitazone	Only 2 patients use this drug
Metformina_pioglitazona	Only 1 patient uses this drug
Acetohexamida	Only 1 patient uses this drug
Tolbutamide	Only 23 patients use this drug
Tolazamide	Only 39 patients use this drug

[a]encounter_id: Identification of a specific hospital visit or patient encounter.

[b]patient_nbr: patient ID number.

[c]payer_code: identifier corresponding to 23 distinct values of payment method (eg, Blue Cross/Blue Shield, Medicare, patient payment).

Table 2. Patient characteristics, description, and their corresponding values.

Variable	Description	Values
Gender	Patient gender (self-identified)	0, 1
Age	Patient age (years)	5, 15, 25, 35, 45,…95
admission_type_id	Identifier corresponding to 8 different types of admissions: emergencies, accidents, newborns, and others	1-8
discharge_disposition_id	Identifier of the discharge type (eg, discharged to home, psychiatric hospital, medical facility)	1-28
admission_source_id	Identifier of the admission source (eg, transfer from hospice, transfer from an ambulatory surgery center)	1-11, 13-14, 20, 22, 25
time_in_hospital	Number of days between admission and discharge	1-14
num_lab_procedures	Number of laboratory tests performed during the encounter	1-132
num_procedures	Number of procedures performed during the encounter	0-6
num_medications	Number of different drugs (generic names) administered during the encounter	1-81
number_outpatient	Number of outpatient visits	0-42
number_emergency	Number of emergency visits	0-76
number_inpatient	Number of inpatient visits	0-21
number_diagnoses	Number of diagnoses	1-16
max_glu_serum	Range of the result of the serum glucose level or if the test was not performed	0, 1
a1cresult	Range of the result of the hemoglobin A1C level or if the test was not performed	0, 1
change	If there is a change in medication	0, 1
diabetesmed	If the patient has been prescribed medication for diabetes	0, 1
readmitted	Days to inpatient readmission; these categories will be relabeled	0, 1, 2
African American, Asian, Caucasian, Hispanic, Other	Patient's race	0, 1
circulatory	If the patient is admitted with circulatory system problems, the variable takes the value of 1	0, 1
diabetes	If the patient is admitted with diabetes-related problems, the variable takes the value of 1	0, 1
digestive	If the patient is admitted with digestive system problems, the variable takes the value of 1	0, 1
genitourinary	If the patient is admitted with genitourinary problems, the variable takes the value of 1	0, 1
injury	If the patient is admitted with injuries, the variable takes the value of 1	0, 1
musculoskeletal	If the patient is admitted with musculoskeletal problems, the variable takes the value of 1	0, 1
neoplasms	If the patient is admitted with neoplasms, the variable takes the value of 1	0, 1
respiratory	If the patient is admitted with respiratory system problems, the variable takes the value of 1	0, 1
other2	If the patient is admitted with other complications, the variable takes the value of 1	0, 1

Table 3. Final drugs used in the data set among 24 total drugs (N=100,000 patients).[a]

Drug	Patients, n (%)
Insulin	54,383 (53.44)
Metformin	19,988 (19.99)
Glipizide	12,686 (12.69)
Glyburide	10,650 (10.65)
Pioglitazone	7328 (7.33)
Rosiglitazone	6365 (6.37)
Glimepiride	5191 (5.19)
Repaglinide	1539 (1.54)
Glyburide-metformin	706 (0.71)
Nateglinide	703 (0.70)

[a]Some patients were administered more than one drug.

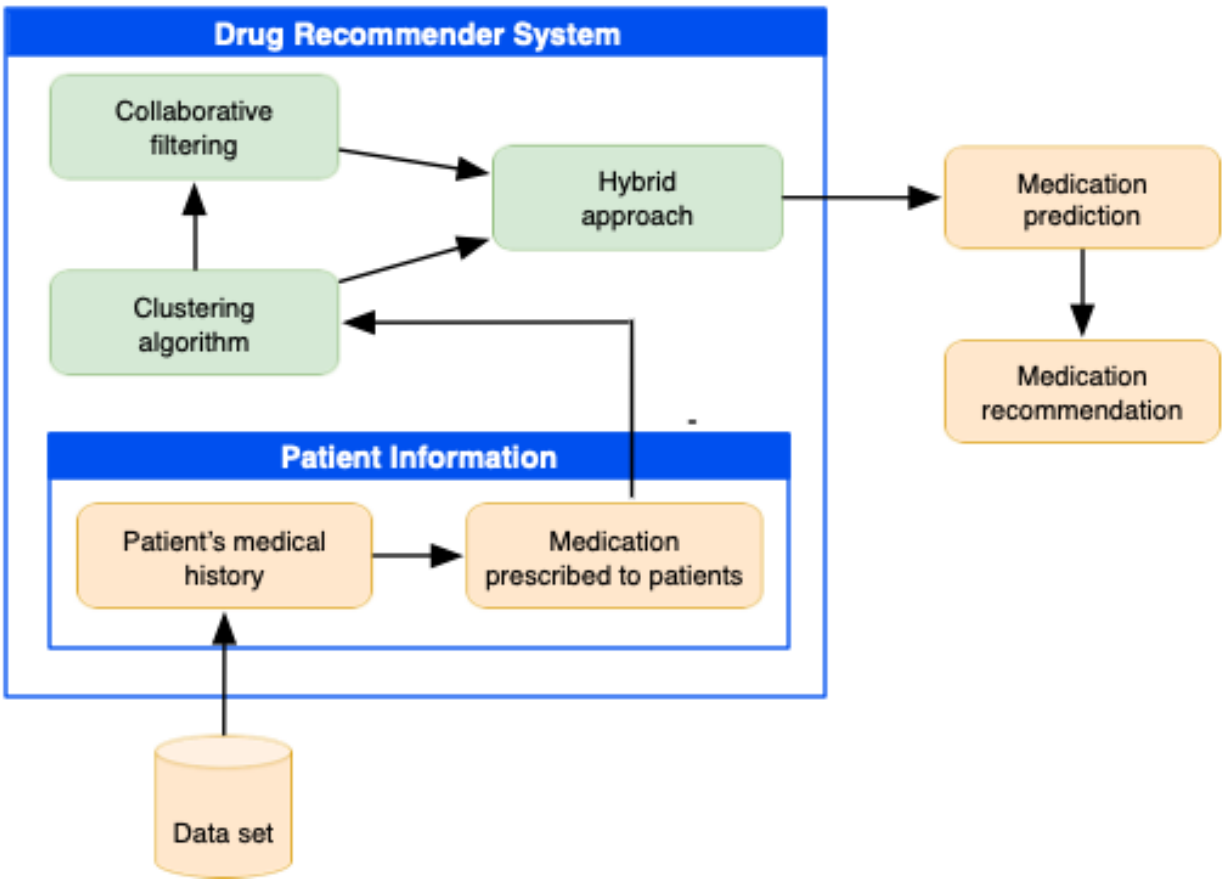
Clustering-Based Recommendation

General Approach

The proposed RS is based on the collaborative filtering approach to represent the drugs prescribed to each patient according to the dose given. The clustering technique was applied to group patients with similar characteristics.

Figure 2 shows a schematic of our proposed method, where the relationship between the elements of the system can be appreciated. The patient's information is first obtained from the data set, such as clinical history, treatments, and the medication prescribed, and then the data are further processed for the application of clustering algorithms. The collaborative filtering technique is then applied to represent the patient's explicit data (user-medication-dose). According to the group to which the patient belongs, the prediction of the medications is made. Finally, the recommendation is made considering the drugs with the highest prediction value.

Figure 2. Schematic of the proposed recommendation approach.



Patient Grouping

To obtain the patient groups, we tested two clustering algorithms to determine the algorithm that provides the best result: the partitional K-means algorithm and the density-based spatial clustering of applications with noise (DBSCAN) algorithm. For this process, it was first necessary to normalize the data and reduce the dimensionality of the data set using principal component analysis (PCA). The optimal number of clusters was determined using the Silhouette coefficient. Patients sharing the same characteristics will be part of the same cluster.

The results of both clustering algorithms were compared using the value of the Silhouette coefficient obtained.

Drug Prediction

For calculation of the prediction performance, we randomly divided the data set into two parts: 80.00% (4142/5177) for training the algorithm and the remaining 20.00% (1035/5177) for testing.

The collaborative filtering approach requires users, items, and ratings; in our case, these elements were replaced by patients, drugs, and drug dosage, respectively, where a value of 1 means that the patient's drug dosage was decreased and a value of 2 means that the drug dosage was increased. With these elements, we proceeded to the construction of the collaborative filtering matrix, as shown in Table 4.

The matrix was completed with the prediction of the drug dose value for the patients in each cluster, which was calculated by applying the cosine similarity measure for each of the clusters.

Table 4. Collaborative filtering matrix.

	Medication 1	Medication 2	Medication 3	Medication 4	Cluster
Patient 1	1	0	2	1	1
Patient 2	0	1	1	1	2
Patient 3	2	0	0	1	1
Patient 4	1	2	0	0	3
Patient 5	1	1	2	0	2

Results

Overall Clustering Results

First, PCA was used to reduce the dimensionality of the data set, resulting in 8 components explaining most of the variance of our data (ie, these components explained 62% of the total variance). We then applied the two clustering algorithms with the 8 principal components. Table 5 shows the best results of the experimentation with the K-means and DBSCAN algorithms.

In comparison with DBSCAN, the K-means algorithm had a higher Silhouette coefficient and a much lower number of clusters for the same data set. Moreover, K-means had a much faster execution time, which means that this algorithm presents lower computational complexity in terms of execution compared with that of DBSCAN. Therefore, the clustering results obtained with K-means were further considered for the calculation of prediction and recommendations. Analysis of the clusters was performed considering the clustering obtained with the K-means algorithm, which showed the best results.

Cluster 4 had the highest number of patients, followed by clusters 2 and 6. Cluster 3 had the smallest number of patients due to the different characteristics considered to group similar patients, such as the diagnosed diseases, race, main drugs administered, and most representative age range.

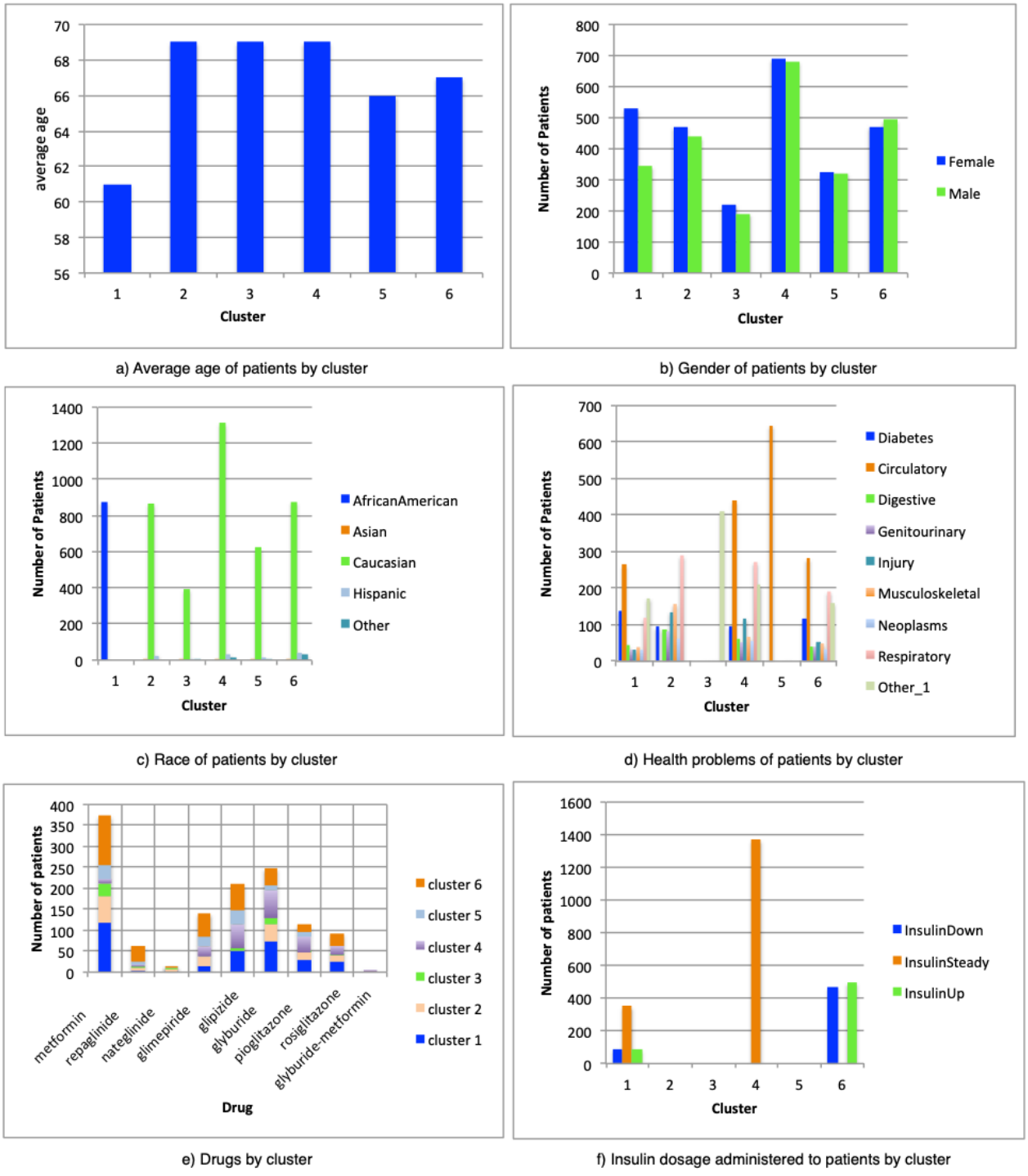
Figure 3 provides details about the six clusters, including the variables analyzed, such as age, gender, race, health problems, medications, and information about the doses of insulin administered to the patient.

Table 5. Comparative analysis of the performance of the algorithms.

Algorithm	Number of clusters	Silhouette coefficient	Execution time
K-means	6	0.654	15 minutes, 24 seconds
DBSCAN[a]	200	0.611	20 minutes, 31 seconds

[a]DBSCAN: density-based spatial clustering of applications with noise.

Figure 3. Details about the six clusters.



Characteristics of Clusters

Cluster 1

Cluster 1 (n=875) included patients who were administered most of the drugs shown in Table 3. In this cluster, 60.6% (n=530) of the patients were female and the remaining 39.4% (n=345) were male. All patients in this group were of African American race. These patients suffered from health problems in the circulatory and respiratory systems directly related to diabetes. The most representative age range in this group was 45-75 years.

Cluster 2

Cluster 2 (n=910) was characterized by a homogeneous distribution of male and female patients; the main diseases diagnosed for these patients involved the respiratory system. Almost all of the patients in this group (n=867, 95.3%) were Caucasian. The main drugs administered for this group were metformin, glipizide, and glyburide, and the dosages were often changed (increased and decreased). Another relevant

characteristic of this cluster is that the patients had not been administered insulin. The most representative age range of patients in this group was 55-85 years.

Cluster 3

In cluster 3 (n=412), the main drugs administered to patients were glyburide, metformin, and glipizide, whereas insulin had also not been administered to this group. There was diversity in the age range of the patients within this cluster, with the most representative range being 55-85 years. The patients in this cluster also tended to be diagnosed with other diseases such as metabolic disease and diseases of the nervous system. All of these patients belonged to the Caucasian race.

Cluster 4

The patients in cluster 4 (n=1369) were mainly administered insulin, metformin, and glipizide; although patients in this group were prescribed other drugs, their proportions were relatively lower. The main diseases diagnosed in this group of patients were diseases of the circulatory and respiratory systems. This cluster grouped the largest number of patients analyzed and the predominant age range was 55-85 years.

Cluster 5

Cluster 5 (n=646) was largely characterized by patients in this group having received more treatments with the following drugs: glyburide, glipizide, and metformin. Insulin had not been administered to this group, and all patients were diagnosed with diseases related to the circulatory system. The most prominent age range was 55-75 years.

Cluster 6

In cluster 6 (n=965), patients were mainly administered the following drugs: insulin, metformin, glipizide, and glyburide. The main diseases diagnosed in this group of patients were those of the circulatory system and diseases related to diabetes mellitus, among others. The predominant age range was 45-75 years.

Quality of Predictions

To evaluate the quality of the predictions, we used the mean squared error (MSE), which penalizes more severely when the error is higher [20]. A significant MSE value of 0.53 was obtained in the test set, which means that the system is capable of obtaining good predictions. As a measure of error, a lower the error value indicates a more efficient RS [21].

Quality of Recommendations

The quality metric is fundamental to measuring the performance of our RS, since it provides information on the proportion of recommended drugs that are relevant for the user. In our case, a drug is considered relevant when the value of the dose is greater than 1. In the experiments, a precision value of 0.61 was obtained, which indicated that the system provides acceptable recommendations.

Generation of Recommendations

Considering a patient "p" with diabetes who belongs to cluster "c," a list of drugs $\{f_1,f_2,...f_n\}$ with the highest prediction score is recommended by the system. An example of the recommendation for two patients in cluster 1 is presented in Table 6 based on a setting of providing the top 3 recommendations.

Table 6 shows that the recommended medications were similar for both patients in cluster 1; however, the order of recommendation varied according to the prediction score obtained for each medication. This is to be expected since these two patients share similar characteristics in terms of clinical information, personal data, and medical treatments stored in the data set.

Table 6. Drug recommendation for two patients with diabetes in cluster 1.

Patient ID	Drugs recommended[a]
36	Insulin, glipizide, metformin
15	Metformin, insulin, glyburide

[a]Drugs are listed in order of preference (highest to lowest prediction score).

Discussion

Principal Results

Our analysis of the UCI Machine Learning Repository showed that most patients with diabetes have circulatory and respiratory problems, followed by metabolic and nervous system problems. Regarding gender, women with diabetes showed more circulatory and respiratory health problems compared to men. It was determined that diabetes manifests differently for individual patients. In addition, in the analysis according to race (Caucasian, African American, Hispanic, and Asian), different categories were identified according to patient and clinical characteristics: (1) older patients with circulatory, respiratory, and other problems; (2) younger patients with digestive, respiratory, and other problems; (3) patients requiring increased insulin doses; and (4) patients prescribed more than one type of medication.

From the data set used, we further observed that diabetes can occur at any age; however, the disease appears to be more common among middle-aged and older people. The analysis showed that the health outcomes related to diabetes are different for each patient, both at the level of control with medication and health complications. Based on these findings, an RS is required to provide support to health care professionals to facilitate the management and control of this chronic disease. Therefore, a cluster-based RS was proposed to help recommend drugs to patients with diabetes. The clustering technique was used to identify groups of patients based on their similar characteristics, and the collaborative filtering technique was used to present information on the doses of the medications

administered to patients. Our experimental results showed acceptable performance of the proposed system.

Limitations and Future Work

The use of explicit information (ratings) from users enables making more precise recommendations; however, according to Wasid and Ali [22], additional effort is required from users when rating an item. Therefore, obtaining useful ratings without additional effort from users is one of the challenges of RSs based on collaborative filtering.

A typical problem of RSs based on collaborative filtering is that data and ratings are often scarce, resulting in a problem of new user and new item cold start. An alternative solution to overcome this problem is to calculate the similarity of the users based on user profiles [13]. For example, if two users are diagnosed with the same disease, they could be considered similar, even if they have not been administered the same drugs during their treatment. Other characteristics such as gender, age, medications administered to the patient, and diseases diagnosed could help classify patients into clusters. Although this study avoided the cold-start (new user) problem by using patient information for clustering and subsequent recommendation, there is a limitation with medications that are entered in the system and have not yet been prescribed to any patient, since the system would have difficulty recommending such medications (ie, new item cold-start problem).

One solution to this problem is to use the metadata of the new items when making recommendations [13]. Combining both user and item information would provide a hybrid approach to address the new user and new item cold-start problem. Therefore, as future work, we propose to (1) extend the recommendation approach using drug information for the prediction and recommendation process, and (2) consider the clustering results of the DBSCAN algorithm for prediction and analyze whether this can improve the quality of recommendations.

Comparison With Prior Work

We found some previous studies related to the topic of RSs in the health domain; however, our proposed approach differs from these previous works by focusing on combining collaborative filtering with clustering techniques to avoid the cold-start (new user) problem.

The experimental results showed that our recommendation approach performs well in terms of offering good predictions and acceptable recommendations considering patient information. In comparison with the study of Sanchez et al [13] who recommended educational content about diabetes, our work focuses on recommending medications to patients with diabetes. Galiano and Paccanaro [23] used collaborative filtering for the prediction of medication side effects to provide recommendations to safety professionals based on a latent factor model. A latent factor method could also be considered for drug prediction to patients with diabetes, and these results can be compared with those obtained using the clustering-based recommendation approach. Consequently, RSs have been developed based on multiple methods that could be combined with clustering techniques to improve the recommendation process. For example, Chung and Jung [24] proposed a knowledge-based cluster model to improve prediction accuracy and make health care recommendations.

Conclusions

We present an RS that is capable of suggesting medications suitable for patients with diabetes. This system considers user metadata to alleviate the cold-start (new user) problem, obtaining groups of patients with similar characteristics using clustering techniques, which are then used to recommend drugs for patients in the same cluster.

To measure the performance of the recommender system, the quality of the predictions and recommendations was evaluated. In the case of prediction accuracy, a significant MSE value was obtained and acceptable accuracy was found in the quality of recommendations, which can be further improved by using information from more drugs or combining with another collaborative filtering approach such as an item-based approach.

The proposed system offers a new method to provide support to health care personnel during the medical care of patients with diabetes by offering recommendations of possible medications that can be considered for the treatment of this disease. In addition, our RS has the advantage of providing recommendations that can be easily explained since the system recommends drugs that have been administered to patients with similar characteristics to the target patient. We believe that this system could be an important tool for health personnel, as it would help to streamline the process of health care and management.

Acknowledgments

The authors would like to thank the Universidad Técnica Particular de Loja for the funding received for the development of this study.

Conflicts of Interest

None declared.

References

1. Lu J, Wu D, Mao M, Wang W, Zhang G. Recommender system application developments: A survey. Decision Support Systems 2015 Jun;74:12-32. [doi: 10.1016/j.dss.2015.03.008]

2. De Choudhury M, Morris MR, White RW. Seeking and sharing health information online: comparing search engines and social media. : Association for Computing Machinery; 2014 Presented at: CHI '14- Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; April 26-May 1, 2014; Toronto, ON p. 1365-1376 URL: https://dl.acm.org/doi/10.1145/2556288.2557214 [doi: 10.1145/2556288.2557214]
3. Huang GJ, Penson DF. Internet health resources and the cancer patient. Cancer Invest 2008 Mar;26(2):202-207. [doi: 10.1080/07357900701566197] [Medline: 18259953]
4. Wong D, Cheung M. Online Health Information Seeking and eHealth Literacy Among Patients Attending a Primary Care Clinic in Hong Kong: A Cross-Sectional Survey. J Med Internet Res 2019 Mar 27;21(3):e10831 [FREE Full text] [doi: 10.2196/10831] [Medline: 30916666]
5. Stellefson M, Shuster J, Chaney B, Paige S, Alber J, Chaney J, et al. Web-based Health Information Seeking and eHealth Literacy among Patients Living with Chronic Obstructive Pulmonary Disease (COPD). Health Commun 2018 Dec;33(12):1410-1424 [FREE Full text] [doi: 10.1080/10410236.2017.1353868] [Medline: 28872905]
6. Zhang Y, Chen M, Huang D, Wu D, Li Y. iDoctor: Personalized and professionalized medical recommendations based on hybrid matrix factorization. Future Generation Computer Systems 2017 Jan;66:30-35. [doi: 10.1016/j.future.2015.12.001]
7. Gujar D, Biyani R, Bramhane T, Bhosale S, Vaidya T. Disease prediction and doctor recommendation system. Int Res J Eng Technol 2018 Mar;5(3):3207-3209 [FREE Full text]
8. Kuanr M, Mohapatra P, Piri J. Health Recommender System for Cervical Cancer Prognosis in Women. 2021 Presented at: 2021 6th International Conference on Inventive Computation Technologies (ICICT); January 20-22, 2021; Coimbatore, India p. 673-679. [doi: 10.1109/ICICT50816.2021.9358540]
9. Princy J, Senith S, Kirubaraj AA, Vijaykumar P. A personalized food recommender system for women considering nutritional information. Int J Pharmaceut Res 2021 Feb 02;13(02). [doi: 10.31838/ijpr/2021.13.02.233]
10. Narducci F, Musto C, Polignano M, de Gemmis M, Lops P, Semeraro G. A recommender system for connecting patients to the right doctors in the HealthNet social network. 2015 Presented at: WWW '15 Companion: Proceedings of the 24th International Conference on World Wide Web; May 18-22, 2015; Florence, Italy p. 81-82. [doi: 10.1145/2740908.2742748]
11. Zhang W, Zou H, Luo L, Liu Q, Wu W, Xiao W. Predicting potential side effects of drugs by recommender methods and ensemble learning. Neurocomputing 2016 Jan;173:979-987. [doi: 10.1016/j.neucom.2015.08.054]
12. Nasiri M, Minaei B, Kiani A. Dynamic recommendation: disease prediction and prevention using recommender system. Int J Basic Sci Med 2016 Jun 29;1(1):13-17. [doi: 10.15171/ijbsm.2016.04]
13. Sanchez Bocanegra CL, Sevillano Ramos JL, Rizo C, Civit A, Fernandez-Luque L. HealthRecSys: a semantic content-based recommender system to complement health videos. BMC Med Inform Decis Mak 2017 May 15;17(1):63 [FREE Full text] [doi: 10.1186/s12911-017-0431-7] [Medline: 28506225]
14. Bianchini D, De Antonellis V, De Franceschi N, Melchiori M. PREFer: a prescription-based food recommender system. Comput Stand Interf 2017 Nov;54:64-75. [doi: 10.1016/j.csi.2016.10.010]
15. Ali S, Amin M, Kim S, Lee S. A hybrid framework for a comprehensive physical activity and diet recommendation system. In: Mokhtari M, Abdulrazak B, Aloulou H, editors. Smart Homes and Health Telematics, Designing a Better Future: Urban Assisted Living. ICOST 2018. Lecture Notes in Computer Science, vol 10898. Cham: Springer; 2018:109.
16. Nowakowska M, Zghebi S, Ashcroft D, Buchan I, Chew-Graham C, Holt T, et al. The comorbidity burden of type 2 diabetes mellitus: patterns, clusters and predictions from a large English primary care cohort. BMC Med 2019 Jul 25;17(1):145 [FREE Full text] [doi: 10.1186/s12916-019-1373-y] [Medline: 31345214]
17. Dong L, Li T, Li L, Wang MZ, Wu Z, Cui J, et al. Clustering patterns of type-type combination in multiple genotypes infections of human papillomavirus in cervical adenocarcinoma. J Med Virol 2019 Nov 07;91(11):2001-2008. [doi: 10.1002/jmv.25553] [Medline: 31347710]
18. Dua D, Graff C. UCI Machine Learning Repository.: Irvine, CA: University of California, School of Information and Computer Science; 2019. URL: https://archive.ics.uci.edu/ml/index.php [accessed 2022-06-27]
19. Strack B, DeShazo JP, Gennings C, Olmo JL, Ventura S, Cios KJ, et al. Impact of HbA1c measurement on hospital readmission rates: analysis of 70,000 clinical database patient records. Biomed Res Int 2014;2014:781670 [FREE Full text] [doi: 10.1155/2014/781670] [Medline: 24804245]
20. Chen M, Liu P. Performance Evaluation of Recommender Systems. Int J Performability Eng 2017;13(8):1246-1256 [FREE Full text] [doi: 10.23940/ijpe.17.08.p7.12461256]
21. Zhu Z, Yan M, Deng X, Gao M. Rating prediction of recommended item based on review deep learning and rating probability matrix factorization. Electronic Commerce Research and Applications 2022 May 09:101160. [doi: 10.1016/j.elerap.2022.101160]
22. Wasid M, Ali R. An improved recommender system based on multi-criteria clustering approach. Procedia Comput Sci 2018;131:93-101. [doi: 10.1016/j.procs.2018.04.190]
23. Galeano D, Paccanaro A. A recommender system approach for predicting drug side effects. 2018 Presented at: 2018 International Joint Conference on Neural Networks (IJCNN); July 8-23, 2018; Rio de Janeiro, Brazil. [doi: 10.1109/ijcnn.2018.8489025]
24. Chung K, Jung H. Knowledge-based dynamic cluster model for healthcare management using a convolutional neural network. Inf Technol Manag 2019 Aug 12;21(1):41-50. [doi: 10.1007/s10799-019-00304-1]

Abbreviations

DBSCAN: density-based spatial clustering of applications with noise
MSE: mean squared error
PCA: principal component analysis
RS: recommender system
UCI: University of California Irvine

Edited by R Kukafka; submitted 11.02.22; peer-reviewed by A Benis, K Uludag; comments to author 05.04.22; revised version received 18.04.22; accepted 29.05.22; published 15.07.22.

Please cite as:
Granda Morales LF, Valdiviezo-Diaz P, Reátegui R, Barba-Guaman L
Drug Recommendation System for Diabetes Using a Collaborative Filtering and Clustering Approach: Development and Performance Evaluation
J Med Internet Res 2022;24(7):e37233
URL: https://www.jmir.org/2022/7/e37233
doi:10.2196/37233
PMID:35838763

Original Paper

The Dose Response Effects of Digital HIV Care Navigation on Mental Health and Viral Suppression Among Young People Living With HIV: Single-Arm, Prospective Study With a Pre-Post Design

Sean Arayasirikul[1,2], PhD; Caitlin M Turner[1], MPH; Dillon Trujillo[3], MPH; Jarett Maycott[1], BA; Erin C Wilson[1,2], DrPH

[1]Center for Public Health Research, San Francisco Department of Public Health, San Francisco, CA, United States

[2]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA, United States

[3]Fielding School of Public Health, University of California, Los Angeles, Los Angeles, CA, United States

Corresponding Author:
Sean Arayasirikul, PhD
Center for Public Health Research
San Francisco Department of Public Health
25 Van Ness Avenue
San Francisco, CA, 94102
United States
Phone: 1 415 554 9000
Email: sean.arayasirikul@gmail.com

Abstract

Background: The HIV epidemic has revealed considerable disparities in health among sexual and gender minorities of color within the Unites States, disproportionately affecting cisgender men who have sex with men (MSM) and trans women. Social inequities further disadvantage those with intersectional identities through homophobia, antitrans discrimination, and racism, shaping not only those at risk for HIV infection but also HIV prevention and care outcomes. Digital interventions have great potential to address barriers and improve HIV care among cisgender MSM and trans women; however, efficacy of digital HIV care interventions vary and need further examination.

Objective: This study assessed the 12-month efficacy of a 6-month digital HIV care navigation intervention among young people living with HIV in San Francisco, California. We examined dose-response relationships among intervention exposure (eg, text messaging), viral suppression, and mental health. Health electronic navigation (eNav) is a 6-month, text message–based, digital HIV care navigation intervention, in which young people living with HIV are connected to their own HIV care navigator through text messaging to improve engagement in HIV primary care.

Methods: This study had a single-arm, prospective, pre-post design. Eligibility criteria for the study included the following: identifying as cisgender MSM or trans women, being between the ages of 18 and 34 years, being newly diagnosed with HIV, or not being engaged or retained in HIV care or having a detectable viral load. We assessed and analyzed sociodemographics, intervention exposure, and HIV care and mental health outcome data for participants who completed the 6-month Health eNav intervention. We assessed all outcomes using generalized estimating equations to account for within-subjects correlation, and marginal effects of texting engagement on all outcomes were calculated over the entire 12-month study period. Finally, we specified an interaction between texting engagement and time to evaluate the effects of texting engagement on outcomes.

Results: Over the entire 12-month period, this study shows that every one-text increase in engagement was associated with an increased odds of undetectable viral load (adjusted odds ratio 1.01, 95% CI 1.00-1.02; P=.03). Mean negative mental health experiences decreased significantly at 12 months compared to baseline for every one-text increase in engagement (coefficient on interaction term 0.97, 95% CI 0.96-0.99; P<.01).

Conclusions: Digital care navigation interventions including Health eNav may be a critical component in the health delivery service system as the digital safety net for those whose social vulnerability is exacerbated in times of crisis, disasters, or global pandemics owing to multiple social inequities. We found that increased engagement in a digital HIV care navigation intervention helped improve viral suppression and mental health—intersecting comorbid conditions—6 months after the intervention concluded. Digital care navigation may be a promising, effective, sustainable, and scalable intervention.

International Registered Report Identifier (IRRID): RR2-10.2196/16406

(J Med Internet Res 2022;24(7):e33990) doi:10.2196/33990

KEYWORDS

HIV; digital HIV care navigation; young people living with HIV; mHealth; mental health; viral suppression; health equity; digital health; equity; effect; suppression; disparity; men who have sex with men; MSM; transgender; sex

Introduction

Background

The HIV epidemic has revealed considerable disparities in health among sexual and gender minorities of color within the Unites States. This is most evident among cisgender men who have sex with men (MSM) and who accounted for 69% of new HIV diagnoses in in 2019 [1]. Transgender—or trans—people are also a high-impact population despite representing 2% of new HIV diagnoses in 2019 [1]. Between 25% and 28% of trans people are HIV-positive [2]. Trans women are overwhelmingly impacted by HIV infection compared to other trans people, representing 93% of HIV diagnoses for trans adults and adolescents [1]. HIV infection impacts those at the intersection of minority race or ethnicity, sexual orientation, and gender identity contributing to grave health disparities among cisgender MSM and trans women. Black or African Americans represent nearly half of HIV prevalence estimates (46%) for trans women, and over one-third (38%) of HIV diagnoses among cisgender MSM [1].

There are more factors at play that contextualize the drastic disparities among trans women and cisgender MSM. For example, in a meta-analysis of HIV prevalence in the US transgender population estimates that 37% of trans women reported having engaged in sex work, 36% reported the use of an illicit substance, only 39.2% reported being employed, and 30.3% of trans women and trans men reported homelessness or unstable housing [2]. These social determinants of health are not only associated with a higher risk of HIV infection but combined with antitrans discrimination and racism, create barriers to care and are associated with suboptimal HIV care outcomes [3].

Digital interventions have great potential to address barriers and improve HIV care among cisgender MSM and trans women. A systematic review of digital interventions found that overall, digital interventions had a positive impact on HIV care outcomes [4]. A Cochrane review found that interventions using mobile phone text messaging at weekly intervals was effective in increasing adherence to HIV treatment or antiretroviral therapy (ART) [5]. Another study with a sample of trans women living with HIV in Los Angeles, California, tested an automated, unidirectional, text messaging intervention and found that it resulted in a number of HIV care outcome improvements including retention to care and ART adherence [6]. Efficacy can vary among digital intervention components and across HIV outcomes with the most successful interventions combining approaches [4]. Finitis et al [7] conducted a systematic review of text messaging interventions among people living with HIV and found that interventions that supported bidirectional communication, occurred less frequently than daily, included personalized message context, and corresponded with participants' ART dosing schedule led to larger effects on ART adherence compared to standard of care controls. This study examines the impact of a multicomponent digital HIV care navigation intervention on mental health and viral suppression among young people living with HIV in San Francisco, California.

Overview of Health eNavigation

Health eNavigation (eNav) is a 6-month text message–based, digital HIV care navigation intervention where young people living with HIV were connected to their own digital HIV care navigator through bidirectional text messaging to improve engagement in HIV primary care. The intervention included delivery of personalized messages and content that addressed the following topics: (1) HIV care navigation, (2) health promotion and education, (3) motivational interviewing (MI), and (4) social support. HIV care navigation guides participants in knowing where, when, and how to access all health and related services, and increases access to appropriate resources (eg, primary medical care, mental health care, housing, insurance and benefits, etc) [8]. Health promotion and education ensures optimal health literacy for all participants by providing information on the biology of HIV, disease management, communication with providers, risk reduction and healthy behavior, and medication adherence. MI is a technique and a style of counseling that can help resolve the ambivalence that prevents patients from realizing their personal goals [9,10]. Social support is provided through establishing an open, nonjudgmental care relationship between participants and their HIV care navigator to be patient-centered and address topics most important to young people living with HIV. Intervention components are described in depth in a prior study [11].

Informed by 2 health services frameworks, Health eNav transforms how HIV care navigation is delivered, seeking to improve health outcomes by amplifying the reach and value of the patient-centered medical home model and the chronic care model with the use of digital technology [12,13]. Health eNav provides participants with personalized engagement to strengthen the provider-patient relationship to eliminate barriers to care, and increase the efficiency and quality of care [12]. Health eNav delivers digital HIV care navigation by providing increased linkages to community resources in a community-driven, cost-effective way, promoting self-management that empowers participants to take an active role in their health, and offering clinical decision support, information sharing, and proactive care in real time [13].

Methods

Ethics Approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki declaration and its later amendments or

comparable ethical standards. This study protocol was approved by the institutional review board at the University of California, San Francisco (IRB #16-19675).

Study Design, Recruitment, and Data Collection

Data for this analysis were obtained from the Health eNav study at San Francisco Department of Public Health (2017-2018). Health eNav was a digital care navigation intervention designed to improve HIV care linkage and retention and subsequent viral suppression among young cisgender MSM and trans women living with HIV. A digital care navigator delivered the intervention via bidirectional text messaging. This is a single-arm, prospective, pre-post design study. Study procedures are described in depth in a prior study [11].

Eligibility criteria for the study included the following: identifying as cisgender MSM or trans women, being between the ages of 18 and 34 years, and being newly diagnosed with HIV or not being engaged or retained in HIV care or having a detectable viral load. Participants were recruited via convenience sampling from 5 clinics and community-based organizations in San Francisco serving young people living with HIV. If eligible, participants met with research staff at study offices within the San Francisco Department of Public Health, where informed consent was obtained. Of 170 individuals screened, 140 were eligible. However, 20 were subsequently lost to follow-up and were not enrolled. This left a final sample of 120 young cisgender MSM or trans women living with HIV.

This analysis examines data collected from computer-assisted self-interview surveys of self-report data administered at baseline, 6 months, and 12 months and intervention exposure data that characterize the number of text messages sent during the 6-month intervention period. Intervention exposure data were collected and exported on the backend of our text messaging platform.

Measures

Demographics

We analyzed the following sociodemographic information: age at interview (in years), gender identity (trans woman vs man), race or ethnicity (non-Hispanic, Latinx American Indian, or Alaska Native; Asian; Black or African American; Multiracial; White; or Hispanic or Latinx), education level (high school or General Educational Development or at least a college education), current living situation (stable vs unstable), income level in the last month (US $0-250, US $251-600, US $601-1300, or ≥US $1301), and incarceration status in the last 6 months.

HIV Care Continuum Outcomes

We assessed 3 key HIV care continuum outcomes at baseline, 6 months, and 12 months of follow-up: whether participants received primary HIV care within the 6 months prior to their study visit, whether participants were currently taking HIV medications at each of these study visits, and whether participants had an undetectable viral load at each of these study visits.

Mental Health

We measured mental health using the mental health subscale of the 12-item Short-Form Health Survey [14]. A composite score was generated by summing responses to items ranging from 0="none of the time" to 5="all of the time" and detailing the frequencies of the following experiences in the last month: feeling calm and peaceful, having a lot of energy, feeling downhearted and blue; and physical health or emotional problems interfering with social activities. The first 2 items in this list were reverse-scored before creating the composite variable.

HIV-Related Stigma

To measure HIV stigma, we used the shortened revised HIV stigma scale tailored for young people living with HIV [15]. A composite score was generated from summing responses (from 0="Strongly disagree" to 4="Strongly agree") to items such as: loss of friends after disclosing HIV status, feeling like a bad person owing to HIV status, or feeling like most people with HIV are rejected when others find out. For both the mental health and stigma composite scores, higher scores denoted a higher number of mental health issues and stigma experiences, respectively.

Intervention Exposure

We measured the number of text messages sent between participants and the digital navigator during the 6-month intervention period, summing the number of text messages. From this, we created the intervention exposure variable, "texting engagement level," defined as the total number of texts sent or received by each participant over the 6-month period. Text message conversations did not comprise preprogrammed, automated, repeated texts; instead, text messages were bidirectional and delivered by an interventionist in conversation with participants using motivational interviewing techniques to have conversations personalized to participants' individual needs. For example, if a participant identified a need for social support to cope with a new HIV diagnosis, the conversation would center that topic. Alternatively, if a participant needed information about health insurance or a health care appointment or medication adherence reminder, the digital HIV care navigator would provide that information or provide follow-up tailored to participants' individual needs. To ensure a baseline level of engagement in the case that participants were not initiating text messages, the digital navigator attempted to start a conversation with participants by sending one text message each week over the duration of the intervention period. The number of text messages ranged from 24 to 467 text messages.

Statistical Analysis

Of the 120 participants in Health eNav, we restricted our analysis to the 60 participants who completed the 6-month digital care intervention. Participants who did not complete the intervention included people who moved out of our jurisdiction, were incarcerated, withdrew from the study, or lost to follow-up during the intervention period. We hypothesized that participants who completed the 6-month digital HIV care navigation intervention represent a different intervention and outcome experience from those who did not. As a result, analyses were

restricted to intervention completers. Additionally, we excluded 4 participants who experienced interruptions in their phone service, lost their phone for a period of time, or deleted the text messaging app, and as a result, text messaging was not possible. The final analytic sample comprised 56 participants.

First, we characterized the entire sample with baseline sociodemographic data. We then described the mean texting engagement level by sociodemographics, HIV care continuum outcomes, mental health composite score (dichotomized into "low mental health issues" or a score of 0 to 10 vs "high mental health issues" or a score of 11 to 20), and HIV-related stigma composite score (dichotomized into "low HIV-related stigma experiences" or a score of 0 to 15 vs "high HIV-related stigma experiences" or a score of 16 to 30). Given the hypothesized difference in intervention effects from baseline to 6 months and then to 12 months, we assessed all outcomes (HIV care continuum, mental health, and HIV-related stigma outcomes) for a 6-month intervention period and 12-month intervention period using generalized estimating equations (GEE) to account for within-subjects correlation. Marginal effects of intervention exposure (or texting engagement) on all outcomes were calculated over the entire 12-month study period using GEE models. Finally, following the logic of differential intervention effects at 6 and 12 months, we specified an interaction between intervention exposure and time point to evaluate the possible effects of a dose response on all 5 outcomes by 6 months and 12 months. All statistical analyses were conducted in Stata 14 [16]. Comparisons producing *P* values less than .05 were considered statistically significant.

Results

Table 1 displays the distribution of baseline sociodemographics, HIV care continuum outcomes, mental health issues, and HIV-related stigma experiences with accompanying mean intervention exposure (or texting engagement). Younger participants, trans women, Hispanic or Latinx participants and those with multiple races or ethnicities, those with less than a college education, those with unstable housing, those with higher incomes, and those who were recently incarcerated had higher mean texting engagement. Those who did not recently receive HIV care, were not currently taking ART, and those with an undetectable viral load also had higher mean texting engagement. Finally, we observed higher mean texting engagement among those who had higher mental health and HIV-related stigma composite scores.

Table 2 shows that the odds of undetectable viral load increased over the initial 6-month intervention period (odds ratio [OR] 2.07, 95% CI 1.04-4.11; *P*<.01) and the 12-month period (OR 2.98, 95% CI 1.11-8.04; *P*=.03). Mean negative mental health decreased over the 6-month period (mean change estimate 0.18, 95% CI 0.05-0.58; *P*<.01), but not over the 12-month period. All other outcome models produced nonsignificant results.

GEE models over the entire 12-month study period (Table 3) showed that every one-text message increase in engagement or intervention exposure was associated with an increased odds of undetectable viral load (adjusted OR 1.01, 95% CI 1.00-1.02; *P*=.03) and a mean increase in HIV-related stigma experiences (mean change estimate 1.03, 95% CI 1.01-1.05; *P*=.02). To better understand the impact on HIV-related stigma, we conducted a post hoc analysis where we stratified by the timing of HIV diagnosis (within the last year vs prior to the last year) given that the recency of HIV diagnosis could serve as a critical period in which young people living with HIV are particularly vulnerable to HIV stigma. Text messaging was associated with a mean increase in HIV stigma only among those who were recently diagnosed (estimate 1.04, 95% CI 1.01-1.07; *P*<.01). Finally, in testing the effect of intervention engagement on outcomes for the 6-month period compared to the 12-month period, we found that mean negative mental health experiences decreased significantly at 12 months compared to baseline for every one-text increase in engagement (coefficient on interaction term 0.97, 95% CI 0.96-0.99; *P*<.01).

Table 1. Sociodemographics, HIV care continuum outcomes, mental health, and HIV-related stigma among young cisgender men who have sex with men and trans women living with HIV who completed the intervention, overall and by texting engagement, Health eNavigation (N=56; 2017-2019).

Sociodemographics	Baseline, n (%)[a]	Texting engagement level, mean (SD)
Age (years)		
18-24	10 (17.86)	141.60 (55.55)
25-36	46 (82.14)	138.65 (66.26)
Gender identity		
Trans woman	8 (14.29)	146.75 (48.21)
Cisgender Man	48 (85.71)	137.92 (66.62)
Race or ethnicity		
Black, non-Hispanic or Latinx	11 (19.64)	139.82 (61.57)
Hispanic or Latinx	14 (25.00)	156.57 (62.65)
Multiple races, non-Hispanic or Latinx	14 (25.00)	155.14 (73.83)
White, non-Hispanic or Latinx	17 (30.36)	111.29 (52.77)
Education level		
High school or General Educational Development or less	21 (37.50)	149.05 (69.21)
Some college or more	35 (62.50)	133.26 (54.42)
Current living situation		
Unstable	35 (62.50)	145.57 (61.31)
Stable	21 (37.50)	128.52 (68.47)
Income in the last month (US $)		
601-1300	13 (23.21)	152.15 (66.45)
251-600	16 (28.57)	125.56 (64.85)
0-250	14 (25.00)	133.79 (53.17)
≥1301	13 (23.21)	148.77 (74.04)
Incarcerated, last 6 months		
Yes	7 (12.50)	146.57 (55.81)
No	49 (87.50)	138.12 (65.55)
HIV care continuum outcomes		
Received primary HIV care, last 6 months		
Yes	50 (89.29)	136.10 (64.01)
No	6 (10.71)	164.83 (59.88)
Currently taking antiretroviral therapy		
Yes	47 (83.93)	135.66 (63.75)
No	8 (14.29)	145.38 (58.73)
Undetectable viral load		
Yes	36 (64.29)	146.33 (63.12)
No	16 (28.57)	109.69 (53.36)
Mental health and HIV-related stigma outcomes		
Mental health composite score (0-20)		
High mental health issues (11-20)	23 (41.07)	149.17 (72.72)
Low mental health issues (0-10)	33 (58.93)	132.21 (57.32)
HIV-related stigma composite score (0-30)		
High stigma experiences (16-30)	17 (30.36)	171.71 (67.07)

Sociodemographics	Baseline, n (%)[a]	Texting engagement level, mean (SD)
Low stigma experiences (0-15)	39 (69.64)	125.00 (57.94)

[a]Percentages calculated out of total number of participants at baseline who completed the intervention and were included in the analysis (N=56), unless otherwise specified.

Table 2. Differences in HIV care continuum, mental health, and HIV-related stigma outcomes at baseline and 6 months for cisgender men who have sex with men and trans women living with HIV who completed the intervention, Health eNavigation (2017-2019).

	Outcomes of generalized estimating equations[a] over time: 6 months compared to baseline		Outcomes of generalized estimating equations[a] over time: 12 months compared to baseline	
	Effect estimate[b] (95% CI)	*P* value	Effect estimate[b] (95% CI)	*P* value
HIV care continuum outcomes				
Received primary HIV care, last 6 months				
No	Reference		Reference	
Yes	3.11 (0.56-17.18)	.19	0.67 (0.21-2.10)	.49
Currently taking antiretroviral therapy				
No	Reference		Reference	
Yes	0.75 (0.35-1.61)	.46	1.38 (0.42-4.55)	.59
Undetectable viral load				
No	Reference		Reference	
Yes	2.07 (1.04-4.11)	.04	2.98 (1.11-8.04)	.03
Mental health and HIV-related stigma outcomes				
Mental health composite score	0.18 (0.05-0.58)	<.01	0.41 (0.14-1.24)	.12
HIV-related stigma composite score	0.29 (0.05-1.75)	.18	0.21 (0.03-1.22)	.08

[a]Five models were created using generalized estimating equations to estimate the effects of each outcome over a 6- and 12-month intervention period. These models produced odds ratios for dichotomous outcomes and prevalence ratios for continuous outcomes.

[b]Odds ratios for dichotomous outcomes; mean change for continuous outcomes.

Table 3. Differences in HIV care continuum, mental health, and HIV stigma outcomes over 12 months by mean texting engagement for cisgender men who have sex with men and trans women living with HIV who completed the intervention, Health eNavigation (2017-2019)[a].

	GEE effects texting engagement over the 12-month study period	
	Adjusted effect estimate[b] (95% CI)	*P* value
HIV care continuum outcomes		
Received primary HIV care, last 6 months		
No	Reference	
Yes	1.00 (0.99-1.00)	.29
Currently taking ART		
No	Reference	
Yes	1.00 (0.99-1.01)	.75
Undetectable viral load		
No	Reference	
Yes	1.01 (1.00-1.02)	.03
Mental health and HIV-related stigma outcomes		
Mental health composite score	1.00 (0.99-1.02)	.61
HIV-related stigma composite score	1.03 (1.01-1.05)	.02

[a]Five models were created using generalized estimating equations to estimate the effects of each outcome over the entire 12-month period. These models produced odds ratios for dichotomous outcomes and prevalence ratios for continuous outcomes.

[b]Odds ratios for dichotomous outcomes; mean change for continuous outcomes.

Discussion

Principal Findings

Our study found evidence of dose-response effects associated with increases in intervention exposure or text message engagement that led to improved odds of undetectable viral load and decreases in negative mental health experiences. While there are studies supporting the application of mobile health (mHealth) approaches to improve HIV care continuum outcomes such as viral suppression, similar advances at the intersection of mental health and HIV care have lagged [17]. A recent study found that HIV medication adherence was negatively associated with poor mental health experiences such as depression, trauma, and adverse childhood experiences [18]. Saberi et al [18] conducted in-depth interviews with 29 participants and found that young people living with HIV preferred digital approaches to mental health service delivery. Our findings strengthen the use of text messaging interventions to support both mental health and HIV care.

We also found that as text messaging increased, HIV stigma experiences also increased. The mean increase in HIV-related stigma experiences associated with increased engagement in text messaging was a surprising finding. Our post hoc analysis found that text messaging was associated with a mean increase in HIV-related stigma only among those who were recently diagnosed (estimate 1.04, 95% CI 1.01-1.07; $P<.01$), suggesting that HIV diagnosis timing might drive the relationship between texting and HIV-related stigma over the study period. We suspect that people who were recently diagnosed either needed more support to process their diagnosis or the changes in their identity related to their diagnosis; participating in an intervention to improve their engagement in HIV care may have brought HIV-related stigma experiences to the surface in order to be addressed as a potential barrier to HIV care engagement in conversations with their digital HIV care navigator [19]. Though studies have called for differentiated service delivery models for key populations experiencing multiple forms of stigma related to HIV and intersectional identities [20], few mHealth interventions have examined the unique needs of young people recently diagnosed with HIV [19].

Limitations and Future Research

Our results should be interpreted with a number of limitations in mind. First, results from our sample of young cisgender MSM and trans women living with HIV in San Francisco may not generalize to other populations. Similarly, since we included only those who completed the 6-month digital care navigation intervention, the findings may only apply to young cisgender MSM and trans women living with HIV who adhere to interventions of this nature. This intervention was focused on changing how HIV care navigation was implemented to include digital methods for young people living with HIV, and owing to our local epidemic at the time of enrollment, this included both cisgender MSM and trans women. We hypothesized that both groups would benefit from participating because the digital navigation participants received was tailored to their individual needs. We did not sample participants to detect differences between these 2 groups. Measurement bias may be in issue as well. Texting engagement, defined as number of texts sent during the digital care navigation component of the intervention, precludes depictions of texting patterns on a day-by-day basis. Texting engagement could have been intermittent as well. However, restricting to those who completed this component

of the intervention insured that texting patterns were likely consistent over the study period. While this intervention did not use standardized, preprogrammed text messages, our training and approach using MI as a client-centered communication framework was standardized and focused on supporting change talk. Selection bias may have played a role in our study as well. Participants who were actively engaged in substance use or encountering acute housing instability may not have had the time or capacity to participate in our intervention study. Finally, given the small sample size, it is possible that some of our analyses were underpowered to detect true effects.

Implications for Future Studies and Conclusions

The COVID-19 pandemic has disrupted the status quo systems of HIV care [21-23], renewing the critical importance of digital interventions in a time of intersecting epidemics [24]. Digital care navigation interventions including Health eNavigation may be a critical component in the health delivery service system as the digital safety net for those whose social vulnerability is exacerbated in times of crisis, disasters, or global pandemics owing to multiple social inequities [25]. We found that increased engagement in a digital HIV care navigation intervention helped improve viral suppression and mental health—intersecting comorbid conditions—6 months after the intervention ended. Digital care navigation may be a promising, sustainable, and scalable intervention for not only making personalized health care more accessible [26], but also serve as a critical link in centering the whole person in a learning health system [27-29].

Acknowledgments

The authors would like to thank all participants in the study. This work was funded by the Health Resources and Services Administration (award H97HA28895). This study's funding sources had no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

Conflicts of Interest

None declared.

References

1. HIV Surveillance Reports. Centers for Disease Control and Prevention. URL: http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html [accessed 2022-07-12]
2. Becasen JS, Denard CL, Mullins MM, Higa DH, Sipe TA. Estimating the Prevalence of HIV and Sexual Behaviors Among the US Transgender Population: A Systematic Review and Meta-Analysis, 2006–2017. Am J Public Health 2019 Jan;109(1):e1-e8. [doi: 10.2105/ajph.2018.304727]
3. Baguso GN, Turner CM, Santos G, Raymond HF, Dawson-Rose C, Lin J, et al. Successes and final challenges along the HIV care continuum with transwomen in San Francisco. J Int AIDS Soc 2019 Apr 29;22(4):e25270 [FREE Full text] [doi: 10.1002/jia2.25270] [Medline: 31037858]
4. Daher J, Vijh R, Linthwaite B, Dave S, Kim J, Dheda K, et al. Do digital innovations for HIV and sexually transmitted infections work? Results from a systematic review (1996-2017). BMJ Open 2017 Nov 03;7(11):e017604 [FREE Full text] [doi: 10.1136/bmjopen-2017-017604] [Medline: 29101138]
5. Horvath T, Azman H, Kennedy GE, Rutherford GW. Mobile phone text messaging for promoting adherence to antiretroviral therapy in patients with HIV infection. Cochrane Database Syst Rev 2012 Mar 14(3):CD009756 [FREE Full text] [doi: 10.1002/14651858.CD009756] [Medline: 22419345]
6. Reback CJ, Fletcher JB, Kisler KA. Text Messaging Improves HIV Care Continuum Outcomes Among Young Adult Trans Women Living with HIV: Text Me, Girl!. AIDS Behav 2021 Sep 24;25(9):3011-3023. [doi: 10.1007/s10461-021-03352-3] [Medline: 34164763]
7. Finitsis DJ, Pellowski JA, Johnson BT. Text message intervention designs to promote adherence to antiretroviral therapy (ART): a meta-analysis of randomized controlled trials. PLoS One 2014 Feb 5;9(2):e88166 [FREE Full text] [doi: 10.1371/journal.pone.0088166] [Medline: 24505411]
8. Higa DH, Marks G, Crepaz N, Liau A, Lyles CM. Interventions to improve retention in HIV primary care: a systematic review of U.S. studies. Curr HIV/AIDS Rep 2012 Dec 21;9(4):313-325 [FREE Full text] [doi: 10.1007/s11904-012-0136-6] [Medline: 22996171]
9. Davidson R. Motivational issues in the treatment of addictive behaviour. In: Edwards G, Dare C, editors. Psychotherapy, Psychological Treatments and the Addictions. Cambridge: Cambridge University Press; 1996:173-188.
10. Miller WR, Rollnick S. Motivational Interviewing (third edition): Helping People Change. New York, NY: The Guilford Press; 2012.
11. Arayasirikul S, Trujillo D, Turner CM, Le V, Wilson EC. Implementing a Digital HIV Care Navigation Intervention (Health eNav): Protocol for a Feasibility Study. JMIR Res Protoc 2019 Nov 8;8(11):e16406. [doi: 10.2196/16406]
12. Jackson GL, Powers BJ, Chatterjee R, Bettger JP, Kemper AR, Hasselblad V, et al. The patient centered medical home. A systematic review. Ann Intern Med 2013 Feb 05;158(3):169-178 [FREE Full text] [doi: 10.7326/0003-4819-158-3-201302050-00579] [Medline: 24779044]

13. Wagner EH, Austin BT, Korff MV. Organizing Care for Patients with Chronic Illness. The Milbank Quarterly 1996;74(4):511. [doi: 10.2307/3350391]
14. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [doi: 10.1097/00005650-199603000-00003] [Medline: 8628042]
15. Wright K, Naar-King S, Lam P, Templin T, Frey M. Stigma scale revised: reliability and validity of a brief measure of stigma for HIV+ youth. J Adolesc Health 2007 Jan;40(1):96-98 [FREE Full text] [doi: 10.1016/j.jadohealth.2006.08.001] [Medline: 17185215]
16. Stata Base Reference Manual (Release 14). College Station, TX: StataCorp LP; 2015.
17. Kempf M, Huang C, Savage R, Safren SA. Technology-Delivered Mental Health Interventions for People Living with HIV/AIDS (PLWHA): a Review of Recent Advances. Curr HIV/AIDS Rep 2015 Dec 9;12(4):472-480 [FREE Full text] [doi: 10.1007/s11904-015-0292-6] [Medline: 26452644]
18. Saberi P, Dawson Rose C, Wootton AR, Ming K, Legnitto D, Jeske M, et al. Use of technology for delivery of mental health and substance use services to youth living with HIV: a mixed-methods perspective. AIDS Care 2020 Aug 28;32(8):931-939 [FREE Full text] [doi: 10.1080/09540121.2019.1622637] [Medline: 31132864]
19. Bayona E, Menacho L, Segura ER, Mburu G, Roman F, Tristan C, et al. The Experiences of Newly Diagnosed Men Who Have Sex with Men Entering the HIV Care Cascade in Lima, Peru, 2015-2016: A Qualitative Analysis of Counselor-Participant Text Message Exchanges. Cyberpsychol Behav Soc Netw 2017 Jun;20(6):389-396 [FREE Full text] [doi: 10.1089/cyber.2016.0435] [Medline: 28622034]
20. Pascoe SJS, Scott NA, Fong RM, Murphy J, Huber AN, Moolla A, et al. "Patients are not the same, so we cannot treat them the same" - A qualitative content analysis of provider, patient and implementer perspectives on differentiated service delivery models for HIV treatment in South Africa. J Int AIDS Soc 2020 Jun 25;23(6):e25544 [FREE Full text] [doi: 10.1002/jia2.25544] [Medline: 32585077]
21. Spinelli MA, Hickey MD, Glidden DV, Nguyen JQ, Oskarsson JJ, Havlir D, et al. Viral suppression rates in a safety-net HIV clinic in San Francisco destabilized during COVID-19. AIDS 2020 Dec 01;34(15):2328-2331 [FREE Full text] [doi: 10.1097/QAD.0000000000002677] [Medline: 32910069]
22. Patel VV, Beil R, Slawek D, Akiyama MJ. HIV Prevention and Treatment in the Context of the COVID-19 in the Bronx, New York: Implications for Practice and Research. AIDSRev 2020 Oct 27;22(3). [doi: 10.24875/aidsrev.20000075]
23. Pinto RM, Park S. COVID-19 Pandemic Disrupts HIV Continuum of Care and Prevention: Implications for Research and Practice Concerning Community-Based Organizations and Frontline Providers. AIDS Behav 2020 Sep 28;24(9):2486-2489 [FREE Full text] [doi: 10.1007/s10461-020-02893-3] [Medline: 32347403]
24. Lesko C, Bengtson AM. HIV and COVID-19: Intersecting Epidemics With Many Unknowns. Am J Epidemiol 2021 Jan 04;190(1):10-16 [FREE Full text] [doi: 10.1093/aje/kwaa158] [Medline: 32696057]
25. Dorn AV, Cooney RE, Sabin ML. COVID-19 exacerbating inequalities in the US. The Lancet 2020 Apr;395(10232):1243-1244. [doi: 10.1016/s0140-6736(20)30893-x]
26. Rodriguez-Villa E, Naslund J, Keshavan M, Patel V, Torous J. Making mental health more accessible in light of COVID-19: Scalable digital health with digital navigators in low and middle-income countries. Asian J Psychiatr 2020 Dec;54:102433. [doi: 10.1016/j.ajp.2020.102433] [Medline: 33271713]
27. Institute of Medicine. Digital Infrastructure for the Learning Health System: The Foundation for Continuous Improvement in Health and Health Care: Workshop Series Summary. Washington, DC: The National Academies Press; 2011.
28. Menear M, Blanchette M, Demers-Payette O, Roy D. A framework for value-creating learning health systems. Health Res Policy Syst 2019 Aug 09;17(1):79 [FREE Full text] [doi: 10.1186/s12961-019-0477-3] [Medline: 31399114]
29. Wagle AA, Isakadze N, Nasir K, Martin SS. Strengthening the Learning Health System in Cardiovascular Disease Prevention: Time to Leverage Big Data and Digital Solutions. Curr Atheroscler Rep 2021 Mar 10;23(5):19. [doi: 10.1007/s11883-021-00916-5] [Medline: 33693992]

Abbreviations

ART: antiretroviral therapy
eNav: electronic navigation
GEE: generalized estimating equation
mHealth: mobile health
MI: motivational interviewing
MSM: men who have sex with men
OR: odds ratio

Edited by T Leung; submitted 02.10.21; peer-reviewed by M Step, D Gerke, J Mitchell; comments to author 23.10.21; revised version received 27.06.22; accepted 08.07.22; published 18.07.22.

Please cite as:
Arayasirikul S, Turner CM, Trujillo D, Maycott J, Wilson EC
The Dose Response Effects of Digital HIV Care Navigation on Mental Health and Viral Suppression Among Young People Living With HIV: Single-Arm, Prospective Study With a Pre-Post Design
J Med Internet Res 2022;24(7):e33990
URL: https://www.jmir.org/2022/7/e33990
doi:10.2196/33990
PMID:35849442

Original Paper

Digital Health Literacy: Bibliometric Analysis

Keng Yang[1,2], PhD; Yekang Hu[3], MA; Hanying Qi[4,5], PhD

[1]Institute of Economics, Tsinghua University, Beijing, China

[2]One Belt-One Road Strategy Institute, Tsinghua University, Beijing, China

[3]China National Health Development Research Center, Beijing, China

[4]The New Type Key Think Tank of Zhejiang Province "Research Institute of Regulation and Public Policy", Zhejiang University of Finance and Economics, Hangzhou, China

[5]China Institute of Regulation Research, Zhejiang University of Finance and Economics, Hangzhou, China

Corresponding Author:
Yekang Hu, MA
China National Health Development Research Center
No. 9 Chegongzhuang Street
Xicheng District
Beijing, 100044
China
Phone: 86 15991750561
Email: freyak5277@hotmail.com

Abstract

Background: Digital health is growing at a rapid pace, and digital health literacy has attracted increasing attention from the academic community.

Objective: The purposes of this study are to conduct a systematic bibliometric analysis on the field of digital health literacy and to understand the research context and trends in this field.

Methods: Methods: A total of 1955 scientific publications were collected from the Web of Science core collection. Institutional co-operation, journal co-citation, theme bursting, keyword co-occurrence, author co-operation, author co-citation, literature co-citation, and references in the field of digital health literacy were analyzed using the VOSviewer and CiteSpace knowledge mapping tools.

Results: The results demonstrate that the United States has the highest number of publications and citations in this field. The University of California System was first in terms of institutional contributions. The *Journal of Medical Internet Research* led in the number of publications, citations, and co-citations. Research areas of highly cited articles in the field of digital health literacy mainly include the definition and scale of health literacy, health literacy and health outcomes, health literacy and the digital divide, and the influencing factors of health literacy.

Conclusions: We summarized research progress in the field of digital health literacy and reveal the context, trends, and trending topics of digital health literacy research through statistical analysis and network visualization. We found that digital health literacy has a significant potential to improve health outcomes, bridge the digital divide, and reduce health inequalities. Our work can serve as a fundamental reference and directional guide for future research in this field.

(J Med Internet Res 2022;24(7):e35816) doi:10.2196/35816

KEYWORDS

digital health literacy; eHealth; digital divide; bibliometrics; VOSviewer; CiteSpace

Introduction

With the development of digital technology, big data, the Internet of Things, artificial intelligence, cloud computing, wearable devices, and so on, digital technologies are constantly being applied to the medical and health fields, giving new vitality to the development of medical health. The World Health Organization (WHO) defines digital health as "the field of knowledge and practice associated with any aspect of adopting digital technology to improve health, from inception to operation"[1]. Digital health expands the concept of eHealth by including other uses of digital technology in health areas such as the Internet of Things, advanced computing, big data analysis, and artificial intelligence [1]. The US Food and Drug

Administration (FDA) defines digital health as having "a broad scope which includes mobile health (mHealth), health information technology, wearable devices, telehealth and telemedicine, and personalized medicine" [2].

As digital health has become more common, digital literacy and health literacy have become important determinants of the usefulness of digital health technologies [3]. Digital literacy refers to "the ability to use information and communication technologies to find, evaluate, create, and communicate information, requiring both cognitive and technical skills" [4]. Health literacy is defined by WHO as "the cognitive and social skills which determine the motivation and ability of individuals to gain access to, understand, and use information in ways which promote and maintain good health" [5]. Based on this, the concept of digital health literacy is gradually taking shape and is receiving more research attention. It has been argued that digital health literacy is a superdeterminant of health that depends on 3 key factors: civic, digital, and health literacy [4]. According to the existing literature, the definition of digital health literacy refers to the application of the definition of health literacy to digital contexts and environments [6]. Therefore, we define digital health literacy in this study as the ability to seek, find, understand, and appraise health information from electronic sources or digital contexts and apply the knowledge gained to addressing or solving a health problem [6,7].

In this study, the concept of digital health literacy also includes eHealth literacy. We argue that the concept of digital health literacy grew from eHealth literacy and was honed thereafter. Norman [8] defined eHealth literacy in 2006 as the ability to seek, discover, understand, and evaluate health information through electronic channels and to apply the knowledge gained to solve health problems. From a definitional perspective, the concepts of digital health literacy and eHealth literacy are similar. Only the channel of accessing and processing health information has changed, evolving from the earlier electronic channel to the digital technology channel. In the existing literature, digital health literacy is often used interchangeably with eHealth literacy [6,9-11]. Specifically, the relationship between digital health literacy and eHealth literacy represents an evolving process of the same thing. With the innovative development of digital technology, eHealth technology is gradually evolving into digital health technology. According to the WHO report, digital health expands and encompasses eHealth [1], and eHealth literacy has become more popular in the era of eHealth technology. With the development of digital health technologies, the term digital health literacy has received increasingly general attention. Taking digital health literacy measurement as an example, early scholars, with Norman [8] at the core, heavily investigated how to measure eHealth literacy. Norman believed that eHealth literacy contains 6 core literacies: traditional literacy, health literacy, information literacy, science literacy, media literacy, and computer literacy. In 2006, Norman developed the eHealth Literacy Scale (eHEALS), which has since been widely used in research on eHealth literacy [12]. The eHEALS is gradually being extended through multinational versions in Italian [13], Spanish [14], German [15], Korean [16], and so on. As research deepens, digital health literacy assessment tools are continually iterated and improved by scholars to enhance the measurement of digital health literacy [6,17,18].

Health literacy is an independent and intermediary determinant of health [19], and improving health literacy helps to improve health outcomes [20,21]. Accordingly, digital health literacy is also a determinant of health [4]. Improving digital health literacy at the population level could address health inequalities, the digital divide [22,23], and public attitudes toward as well as practices and awareness of digital health [1]. That is, improving digital health literacy is a useful solution to emerging health challenges. This has become especially clear during the COVID-19 pandemic, with digital health literacy being a key capability for finding information on COVID-19 on the internet [24]. Therefore, it is important to understand the current situation and trends in digital health literacy research and to identify future research opportunities.

Moreover, there is some bibliometric research related to health literacy and eHealth literacy [25-27]. However, although these studies have done systematic bibliometrics, their core concepts are limited to eHealth literacy. They were unable to comprehensively embrace cutting-edge developments in the field of digital health literacy. In addition to including eHealth literacy, we further encompassed literature on the intersection of digital technology and health literacy and that of digital health and health literacy to expand and enrich the existing research.

Against this background, we use bibliometric analysis and knowledge network visualization to analyze institutional co-operation, journal co-citation, topic bursting, keyword co-occurrence, author co-operation, author co-citation, reference co-citation, and bursting in the field of digital health literacy. In this way, we map the knowledge structure of research in this field, along with research trends and trending topics. To the best of our knowledge, this is the first comprehensive bibliometric analysis in digital health literacy, and this study can provide basic support and directional guidance for future research in this field.

Methods

Data Collection

The Web of Science (WoS) Core Collection served as the data source in this study. This database includes journals that are indexed in the Social Science Citation Index (SSCI), the Science Citation Index Extension (SCIE), and the Art and Humanities Citation Index (A&HCI). The retrieval method relied on the steps outlined by Chen [28]. We retrieved articles containing both the keywords "digital health" and "literacy" as well as those with the keyword "health literacy" and keywords related to digital technology. Our specific screening strategies are shown in Figure 1 and outlined as follows.

First, we retrieved articles with the keyword "digital health" and derivative keywords of the topic. The query command for #1 was TS = ("digital health" OR "digital health care" OR "digital medicine" or "eHealth" OR "eHealth care" OR "eHealth care" OR "e-medicine" OR "telehealth" OR "tele-health" OR "telehealthcare" OR "tele-healthcare" OR "telemedicine" OR "tele-medicine" OR "mHealth" OR "m-health" OR

"mHealthcare" OR "m-healthcare" OR "mobile health" OR "mobile healthcare" OR "mobile medicine" OR "online health" OR "online healthcare" OR "online medicine"). TS refers to the topic tag for search string retrievals.

Second, we retrieved articles with the keyword "literacy." The query command for #2 was TS = ("literacy").

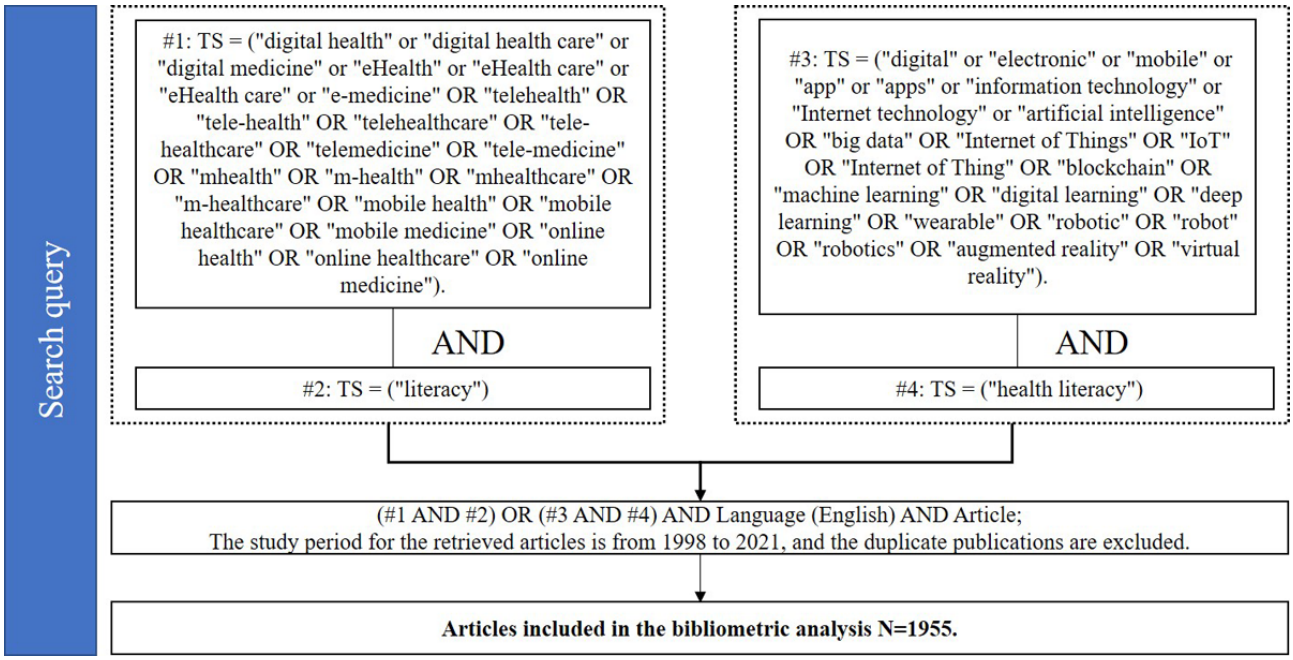
Third, we retrieved articles with digital technology–related keywords. The query command for #3 was TS = ("digital" OR "electronic" OR "mobile" OR "app" OR "apps" OR "information technology" OR "Internet technology" OR "artificial intelligence" OR "big data" OR "Internet of Things" OR "IoT" OR "Internet of Thing" OR "blockchain" OR "machine learning" OR "digital learning" OR "deep learning" OR "wearable" OR "robotic" OR "robot" OR "robotics" OR "augmented reality" OR "virtual reality").

Fourth, we retrieved articles with the keyword "health literacy." The query command for #4 was TS = ("health literacy").

Fifth, we combined the commands above[(#1 AND #2) OR (#3 AND #4)] AND Language (English) AND Article. This was the last step of the screening process. The query command (#1 AND #2) indicates retrieving papers related to both digital health and literacy. This refers to articles that study literacy in the field of digital health. The second query command for the last step (#3 AND #4) indicates retrieving articles related to both digital technology and health literacy. This refers to articles that focused on digital technology in the field of health literacy. Therefore, the combined command of the last step aimed to retrieve articles in English related to "digital health and literacy" or "digital technology and health literacy."

A total of 1955 articles in English were retrieved as research samples. The beginning time point for data collection was 1990, and the study period for the retrieved articles was from 1998 to 2021. The data retrieval took place on September 30, 2021.

Figure 1. The flowchart for data collection.



Analysis Methodologies

bibliometric analysis, or bibliometrics, was first described by Pritchard [29] as "the application of mathematical and statistical methods to articles and other forms of communication." Bibliometrics is a method of information analysis that measures research trends and knowledge structures in a field to obtain quantifiable, objective data [30]. It has been extensively used to quantitative analyze academic literature to describe trending topics and contributions of scholars, journals, and countries [31] and help researchers understand the current research trends, distribution, and core topics in a given field [32].

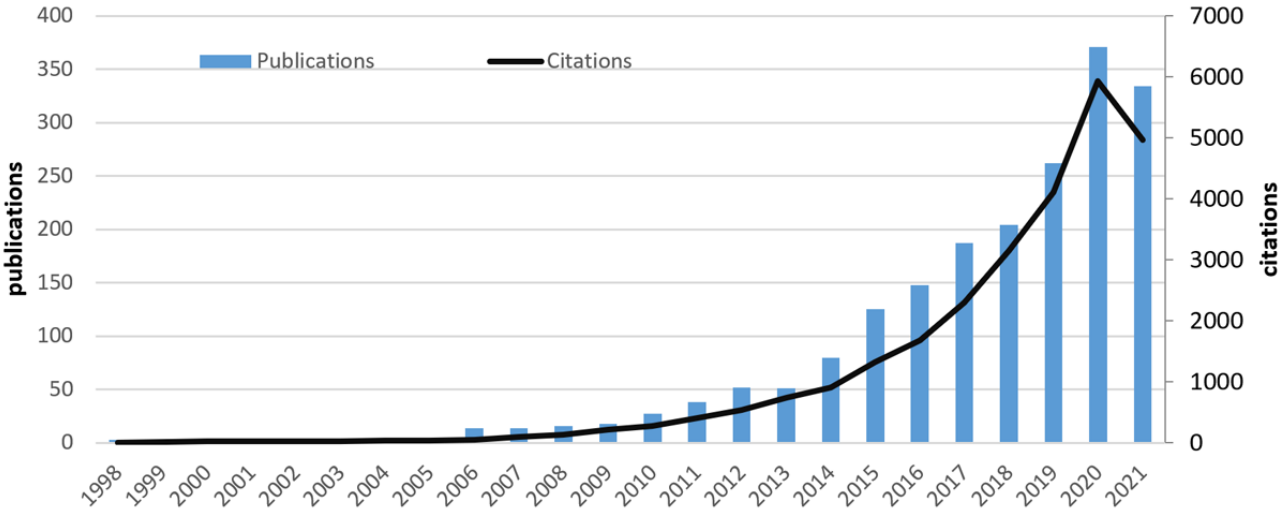
We used scientific mapping tools for bibliometric analysis. Currently popular tools include VOSviewer [33], CiteSpace [34], BibExcel [35], HistCite [35], and others. We selected VOSviewer and CiteSpace as our analysis tools. The reason for this was that VOSviewer has better visualization in network and cluster analysis, and CiteSpace is better in literature timeline analysis. A combination of these two tools can better achieve our research goals. VOSviewer was developed by Van Eck and Waltman [33] and features a powerful bibliometric maps function that can clearly visualize the network of literature, keywords, authors, and so on. Using VOSviewer (version 1.6.16), we drew diagrams for institutional co-operation, journal co-citation, keyword co-occurrence, author co-operation, author co-citation, and literature co-citation. CiteSpace was developed by Chen [34] to implement 2 complementary visualizations: cluster view and time zone view. We used CiteSpace (5.8.R1) to draw 2 analysis mappings: theme bursting and reference bursting. For the distribution of publications, countries, institutions, journals, and authors, Microsoft Excel was employed to perform the analyses.

Results

Overall Trends in Publications and Citations

Figure 2 depicts trends in publications and citations in digital health literacy. Our search yielded 1955 articles covering the period from 1998 to 2021. From when the first article was published in 1998 until 2021, the number of articles published continuously increased, with an obvious overall growth trend. In terms of average annual publications, the trend can be divided into 3 stages: (1) 1998 to 2005 was the gestation period, and the number of publications stayed in the single digits, with an average of 3 articles per year; (2) 2006 to 2013 was a slow growth period, and the number of publications remained below 100 each year. The average number of publications was about 34 annually; (3) Since 2014, digital health literacy research has undergone rapid growth. As of September 30, 2021, the average annual number of articles was 233, with a peak in 2020 (n=392, 20.05%). Publications increased dramatically from 2014 to 2021, accounting for 87.5% (1711/1955) of all publications, and are still increasing. Citations have a similar overall development trend.

Figure 2. Total publications and citations from 1998 to 2021.



Country Distribution

About 91.7% (1792/1955) of the publications were from the top 10 productive countries or regions shown in Table 1. The United States ranked first with 913 published articles, accounting for 46.7% of all 1955 publications and far exceeding the numbers of other countries. Australia was second (198/1955, 10.1%), followed by the United Kingdom (136/1955, 7%), Canada (108/1955, 5.5%), China (104/1955, 5.3%), and Germany (102/1955, 5.2%).

Table 1. Top 10 publication countries/regions (N=1955).

Countries/regions	Publications, n (%)
United States	913 (47)
Australia	198 (10.1)
England	136 (7)
Canada	108 (5.5)
China	104 (5.3)
Germany	102 (5.2)
Netherlands	99 (5.1)
Switzerland	45 (2.3)
Denmark	44 (2.3)
Sweden	43 (2.2)

Institution Distribution

The top 10 research institutions according to publication number are presented in Table 2. The institution that published the most papers at 100 publications was the University of California System, accounting for 5.12% (n=1955) of all publications. Harvard University was second, with 64 (3.3 %) publications, and the State University System of Florida came in third, with 3.2% (n=1955) of all publications.

Of the top 10 institutions, 1 is a government agency: the US Department of Veterans Affairs, which is the administrative department for veterans and mainly focuses on health literacy among veterans. The other 9 institutions are all universities.

We also analyzed coauthorship among institutions. A total of 959 institutions with high association strength were automatically identified using VOSviewer and were used to draw the institutional co-operation network shown in Figures 3 and 4. A total of 31 clusters were formed, and different colors represent different clusters. In these figures, node size refers to publications, circle color to clustering, and link thickness to co-operation strength. The minimum number of documents of an institution is 1 publication. The minimum number of citations of an institution is 0. The largest cluster has 80 institutions, and the smallest has 7. We clustered the institutions network using the default VOS clustering, which uses a clustering algorithm similar to modularity-based clustering. Of these clusters, 4 stand out in comparison to others, and these co-operation networks have obvious geographical characteristics. In Figure 3 (see Multimedia Appendix 1 for full-size image), the red cluster is a collaborative network with the University of California System at its core, while the green cluster is a collaborative network with Harvard University, the State University System of Florida, and the US Department of Veterans Affairs as its core. Institutions in these 2 clusters are primarily from the United States. The thick line between the University of California System and Harvard University suggests a high level of collaboration. The orange cluster is a collaborative network with the University of Sydney, University of Melbourne, and University of Queensland as the core. These institutions are mainly from Australia and the Commonwealth. The main institutions in the blue cluster (lower right corner) are largely based in emerging market countries such as Hungary, Turkey, and Vietnam, and they mainly co-operate with other institutions also located in emerging markets.

Figure 4 shows the average publication year of articles published by each institution (See Multimedia Appendix 2 for full-size image). Time is represented by different colors. The darker the color, the earlier the average publication year of the institution. As seen, the average publishing year of institutions from the United States was earlier, followed by institutions from Australia. The average publishing year of institutions from emerging markets was closer to 2021; this indicates that research attention to digital health literacy has gradually spread from high-income countries to emerging market countries and regions.

Table 2. Top 10 institutions of publications.

Institution	Publications, n (%)
University of California System	100 (5.1)
Harvard University	64 (3.3)
State University System of Florida	63 (3.2)
University of North Carolina	49 (2.5)
University of Sydney	49 (2.5)
US Department of Veterans Affairs	49 (2.5)
University of Texas System	44 (2.3)
Northwestern University	42 (2.2)
University System of Maryland	37 (1.9)
University of London	33 (1.7)

Figure 3. Institutional coauthorship network (1998-2021).

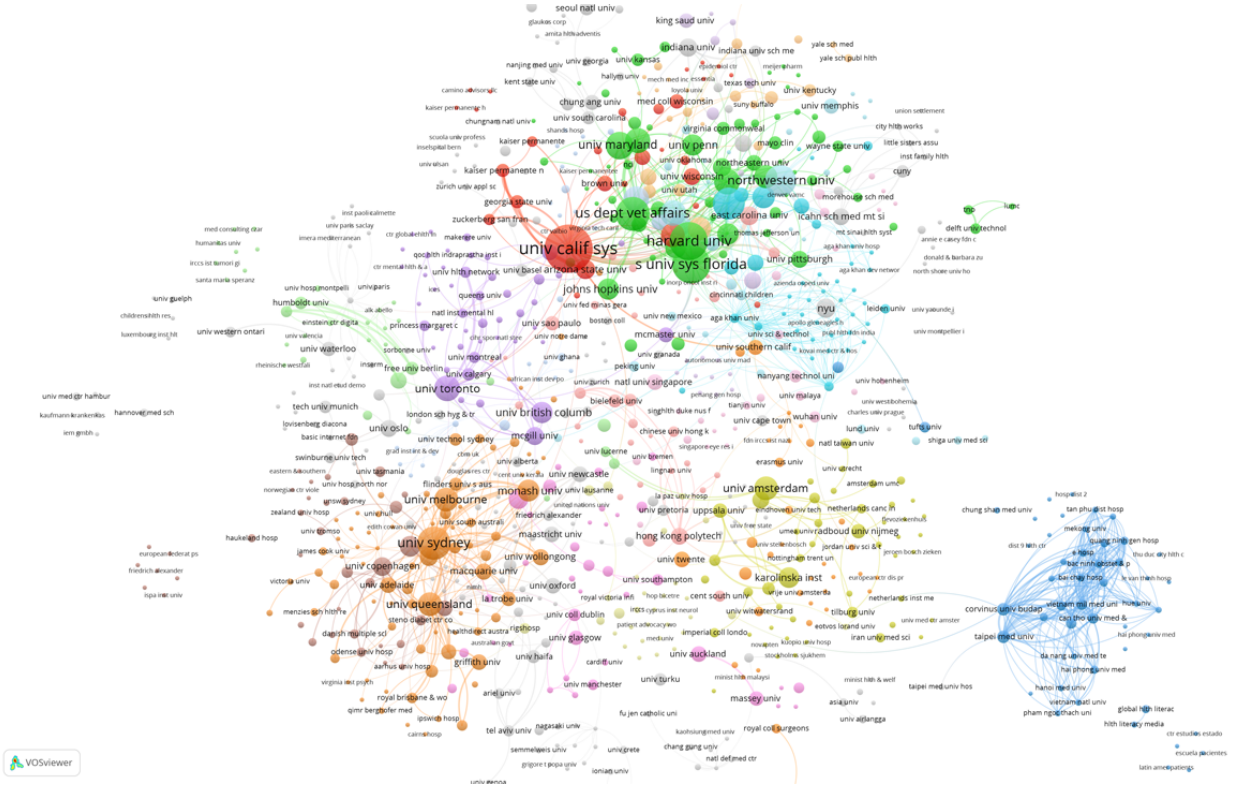
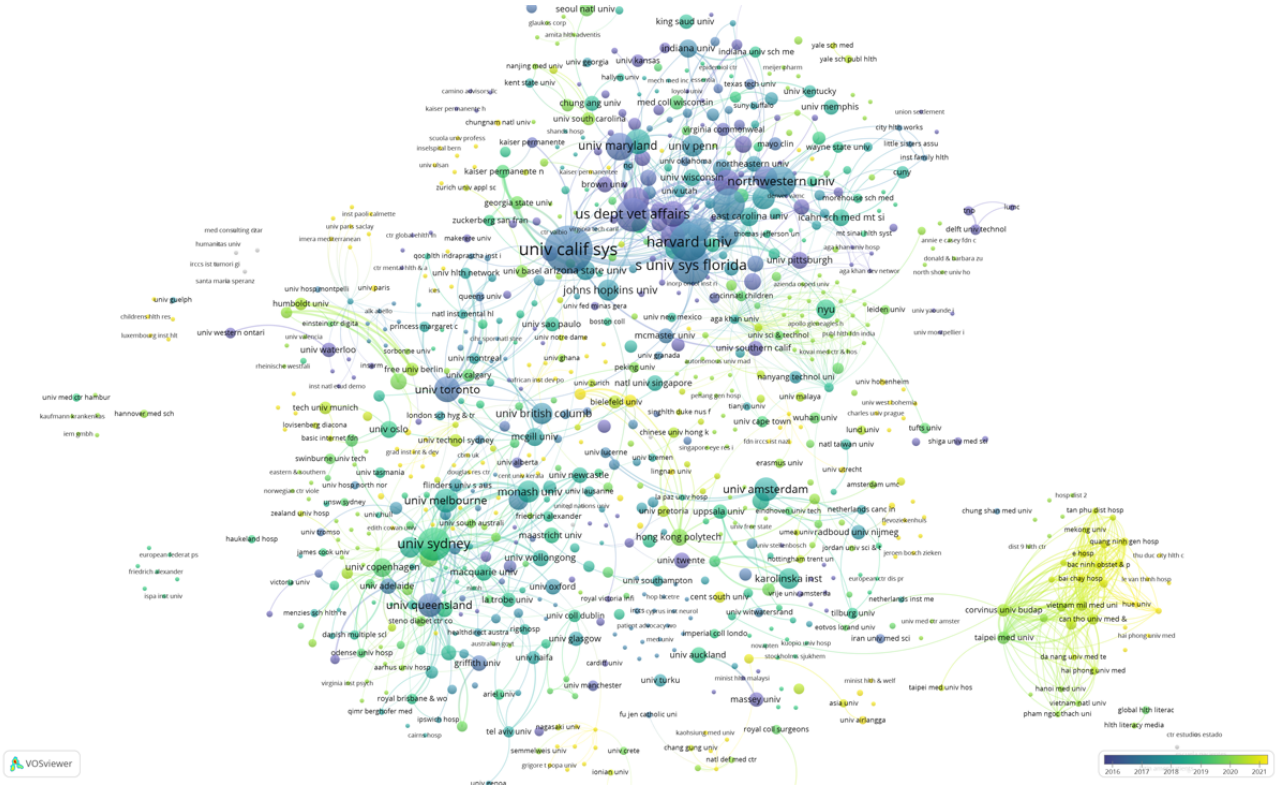


Figure 4. Time trend map of organizations' coauthorship.



Journal Distribution

The 1955 citing articles in this paper were published in a total of 710 different journals. Table 3 shows the top 10 journals in terms of publication number and reports both their total citation frequency and average citation frequency. The *Journal of Medical Internet Research* ranked first with 207 publications, accounting for 10.6% of all 1955 articles, and it had the highest citation frequency at 6663 citations, accounting for 24.7% of all articles' total citations (N=27,012). Two other journals from the same publisher (JMIR Publications) also ranked highly in terms of publications, namely *JMIR mHealth and uHealth* (62/1955, 3.2%) and *JMIR Research Protocols* (39/1955, 2%),

indicating that JMIR Publications paid more attention to digital health literacy research within the study period.

Table 3 shows that when journals are ranked according to articles published, their citation numbers are not consistent with their ranking. Some journals have relatively high citation numbers even though they have published relatively few articles. Average citation numbers illustrate that certain journals have higher average citation numbers than others, which may reflect the relative quality of these journals' publications.

Table 4 presents the top 10 journals ranked by citation number. As the table indicates, some journals have higher citation numbers despite having fewer publications. For example, the Journal of General Internal Medicine has the second highest number of citations (n=894, 3.3%) but only published 9 papers with an average of 99.33 citations, indicating that the papers published in this journal have played a significant role in promoting digital health literacy research. Annals of Internal Medicine had the highest average citation numbers (n=155, 0.58%) with only 3 publications; the total number of citations for this journal was 465(1.7%), indicating that the 3 publications were relatively important and made significant contributions to digital health literacy research.

Table 3. Top 10 most productive journals.

Journals	Publications[a], n (%)	Citations[b], n (%)	Average citations, n
Journal of Medical Internet Research	207 (10.6)	6663 (24.7)	32.19
JMIR mHealth and uHealth	62 (3.2)	567 (2.1)	9.15
International Journal of Environmental Research and Public Health	46 (2.4)	218 (0.8)	4.74
JMIR Research Protocols	39 (2)	187 (0.7)	4.79
Patient Education and Counseling	37 (1.9)	679 (2.5)	18.35
BMC Public Health	31 (1.6)	536 (2)	17.29
Telemedicine and e-Health	29 (1.5)	404 (1.5)	13.93
BMC Medical Informatics and Decision Making	26 (1.3)	372 (1.4)	14.31
BMJ Open	26 (1.3)	130 (0.5)	5
Journal of Health Communication	24 (1.2)	612 (2.3)	25.50

[a]N=1955.

[b]N=27,012.

Table 4. Top 10 journals according to number of citations.

Journal	Citations[a], n (%)	Publications[b], n (%)	Average references, n
Journal of Medical Internet Research	6663 (24.7)	207 (10.6)	32.19
Journal of General Internal Medicine	894 (3.3)	9 (0.5)	99.33
Journal of the American Medical Informatics Association	702 (2.6)	19 (1.0)	36.95
Patient Education and Counseling	679 (2.5)	37 (1.9)	18.35
Journal of Health Communication	612 (2.3)	24 (1.2)	25.50
JMIR mHealth and uHealth	567 (2.1)	62 (3.7)	9.15
BMC Public Health	536 (2.0)	31 (1.6)	17.29
BMC Health Services Research	532 (2.0)	18 (0.9)	29.56
Annals of Internal Medicine	465 (1.7)	3 (0.2)	155
Telemedicine and e-Health	404 (1.5)	29 (1.5)	13.93

[a]N=27,012.

[b]N=1955.

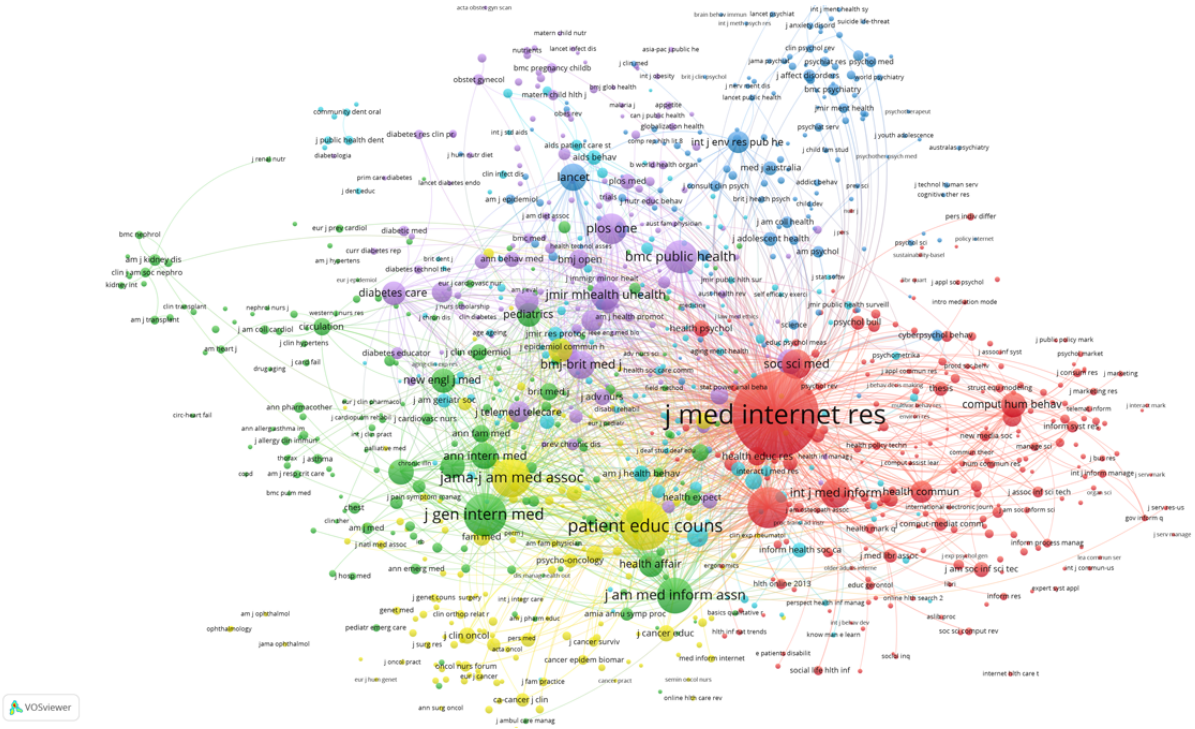
We conducted a co-citation analysis of journals based on the references of 1955 articles. We used VOSviewer to form a journal co-citation network with 6 clusters, as presented in Figure 5. In this figure, color represents cluster, circle size represents the outgoing document quantity, and line thickness represents co-citation intensity (See Multimedia Appendix 3 for full-size image). The co-citation network consists of 1000 journals, and each journal has more than 10 citations and strong co-citation correlation. As indicated by the largest red circle in the figure, the *Journal of Medical Internet Research* is at the core of the whole co-citation network, and it has the largest number of citations and greatest co-citation intensity. It is a

leading journal in digital health literacy research and focuses mainly on fields such as health informatics and digital health. In addition, the *Journal of Health Communication*, which features in the same red cluster, is also important, indicating that their co-citation relationships are relatively strong. This journal also focuses on health informatics.

In the yellow cluster, *JAMA: Journal of the American Medical Association* is important. It is a top journal in the field of medicine and is the most widely circulated general medical journal in the world. *Patient Education and Counseling*, which focuses mainly on patient education and health promotion, is also important in the yellow cluster.

The most important journals in the green cluster are *Journal of General Internal Medicine* and *Journal of the American Medical Informatics Association*. Meanwhile, *BMC Public Health* and *Public Library of Science (PLOS) One* stand out in the purple cluster, and *The Lancet* leads in the blue cluster. All these journals have played an important supporting role in digital health literacy research.

Figure 5. The co-citation network of journals.



Research Topics

We investigated the research topics of digital health literacy in terms of both research categories and research themes. Table 5 shows the top 10 categories of digital health literacy publications. The unit of measurement in the table is the number of publications. The data for these research categories are generated by the WoS database system based on search results. Health care sciences services (569/1955, 29.1%), medical informatics (436/1955, 22.3%), and public environmental occupational health (427/1955, 21.8%) were the 3 most important research categories for digital health literacy. In addition, digital health literacy also covers "health policy services," "information science library science," "nursing," "medicine general internal," "oncology," "computer science information systems," and "communication."

To better understand the time trends of research topics, we used CiteSpace to analyze theme term bursts. The bursting of a term refers to when the citation strength of a term suddenly grows, which can reflect that the term has attracted more attention than before. The burst strength is calculated by the default Kleinberg algorithm of CiteSpace. The theme terms were extracted from the titles and abstracts of the citing articles. The top 20 terms in each year were selected to construct a co-occurrence network of theme terms. On this basis, the top 46 theme terms with strong burstiness were extracted, which were all the burst items found by CiteSpace. The list of these terms was ranked by the starting year of bursts. Figure 6 reports the top 25 strongest theme terms, and the list of all 46 theme terms citation bursts is reported in Multimedia Appendix 4. In Figure 6, "terms" in the first column refers to the theme terms, "year" in the second column refers to the year in which a theme term first occurred, "strength" refers to the citation bursts strength of theme terms, "begin" refers to the starting year in which a theme term burst, and "end" refers to the end year of a bursting theme term.

The theme terms related to the digital health literacy have the following features. Early studies mainly used theme terms that included "electronic medical record" (which was widely cited from 2011 to 2019) and "electronic health record" (which was widely cited from 2013 to 2018); this indicates that studies mainly focused on health records and health information before 2014, which was the gestation period of digital health literacy. The second burst peak appeared in 2014 with "health literacy," and the strength of this burst continues today. In addition, "health information" and "online health information" also began to burst and grow in strength at this stage. Theme terms related

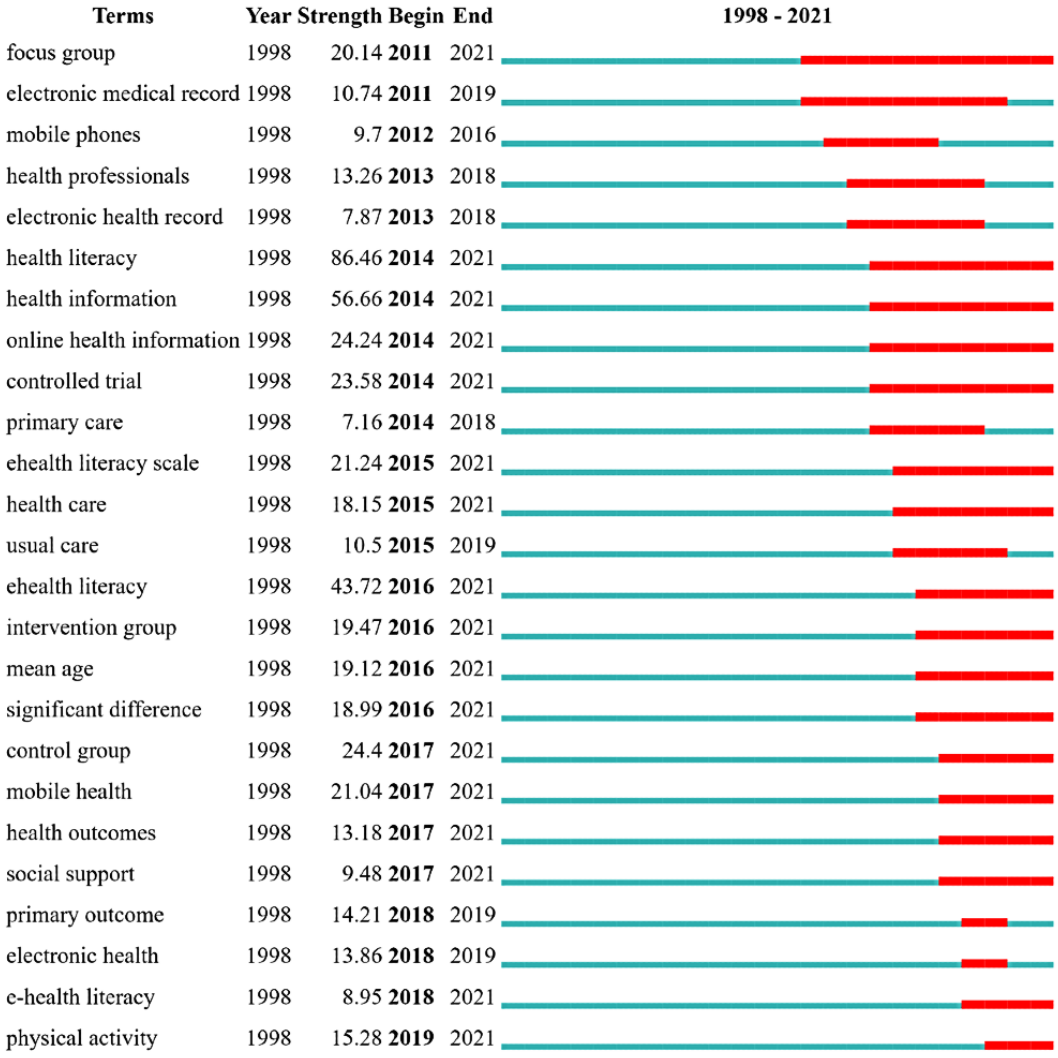
to “health literacy” and “health information” at this stage have remained trending topics. Since 2015, the “eHealth Literacy Scale” has gained popularity. From 2016 to the present, theme terms such as “ehealth literacy” and “electronic literacy” have become popular. “eHealth literacy” has become a trending research topic. Thus, the concept of digital health literacy has gradually matured, and relevant research has exploded.

Table 5. Top 10 research categories.

WoS[a] categories	Publications, n (%)
Health care sciences services	569 (29.1)
Medical informatics	436 (22.3)
Public environmental occupational health	427 (21.8)
Health policy services	143 (7.3)
Information science library science	127 (6.4)
Nursing	116 (5.9)
Medicine general internal	111 (5.6)
Oncology	78 (3.9)
Computer science information systems	76 (3.8)
Communication	69 (3.5)

[a]WoS: Web of Science.

Figure 6. Top 25 terms with the strongest citation bursts.



Co-occurrence of Keywords

The aforementioned research topics analysis obtained theme terms using the titles and abstracts of the surveyed articles. In this section, we analyze the co-occurrence network of keywords. Keywords can indirectly reveal trends and changes in research topics, which is crucial to understanding their development [36]. We used VOSviewer to construct a keyword co-occurrence network (Figure 7). Additionally, Table 6 shows the top 25 keywords according to occurrence. The most frequently occurring keyword was “health literacy,” which appears 666 times and has 580 links, with a total link strength of 4500. In addition, “internet,” “literacy,” “care,” and “eHealth literacy” are also frequently used keywords that rank highly in both occurrence frequency and connection strength.

To better elucidate the co-occurrence relationship between digital health literacy keywords and changing trends, VOSviewer was used to identify a total of 612 different keywords with 5 or more occurrences. Figures 7 and 8 represent the network and overlay visualization of keyword co-occurrence, respectively. There are 4 clusters in both figures. These clusters were organized by using the default VOS clustering, which is a clustering algorithm similar to modularity-based clustering. In Figure 7, color represents cluster, node size refers to frequency of appearance, and link thickness represents co-occurrence intensity (See Multimedia Appendix 6 for full-size image).

Figure 7. Keyword co-occurrence network.

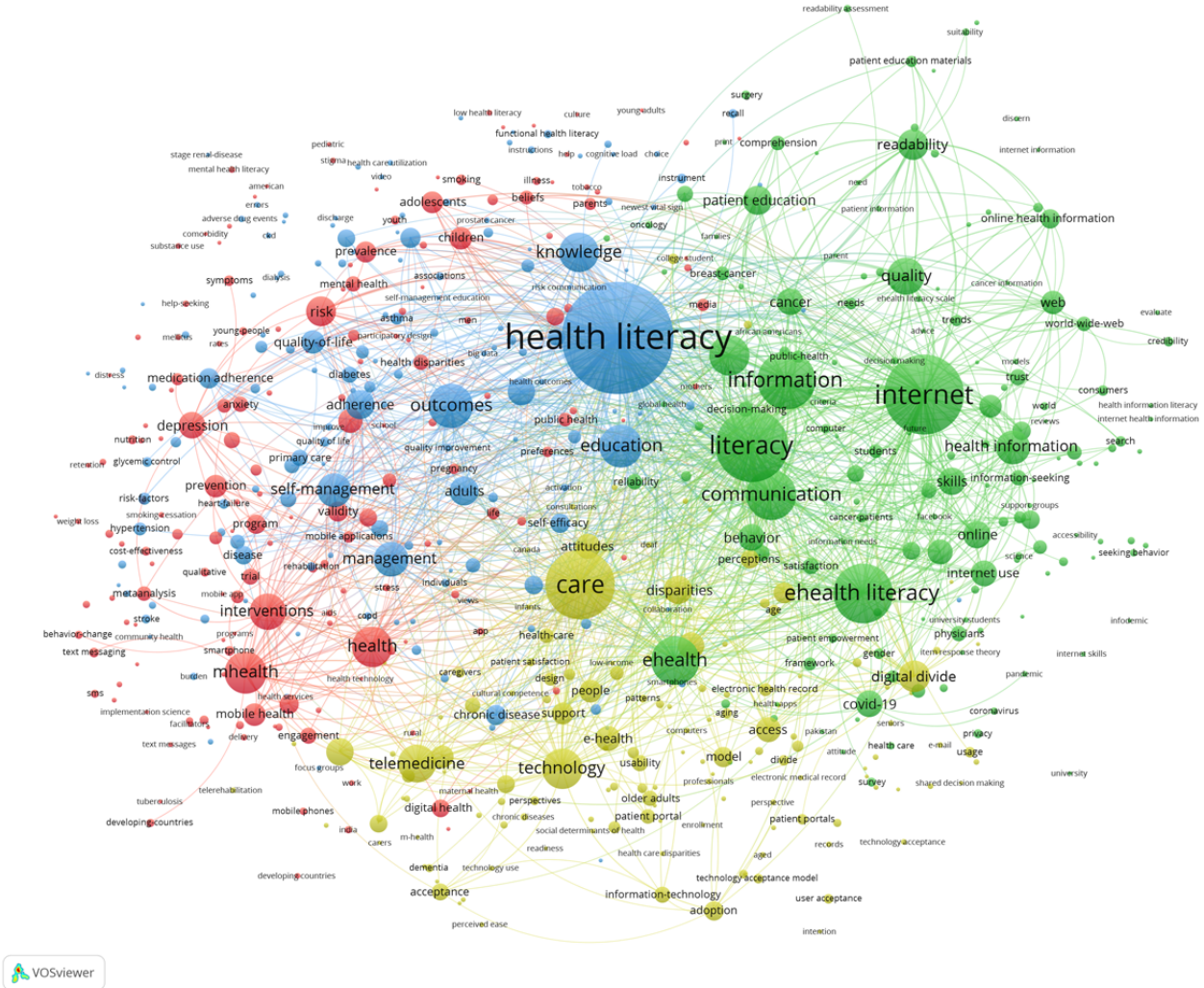


Table 6. Top 25 most frequently occurring keywords of digital health literacy.

Keyword	Occurrences[a], n (%)	Links[b,c], n (%)	Total link strength
Health literacy	666 (7.9)	580 (2.5)	4500
Internet	397 (4.7)	495 (2.1)	2973
Literacy	339 (4.0)	491 (2.1)	2587
Care	316 (3.7)	498 (2.2)	2416
eHealth literacy	252 (3.0)	422 (1.8)	1867
Information	241 (2.9)	432 (1.9)	1903
Communication	194 (2.3)	409 (1.8)	1625
eHealth	186 (2.2)	408 (1.7)	1443
Outcomes	159 (1.9)	374 (1.6)	1274
Education	152 (1.8)	355 (1.5)	1143
mHealth	148 (1.8)	346 (1.5)	1095
Health	141 (1.7)	340 (1.5)	955
Impact	139 (1.6)	361 (1.6)	1085
Technology	138 (1.6)	345 (1.5)	1153
Knowledge	134 (1.6)	330 (1.4)	1029
Telemedicine	127 (1.5)	306 (1.3)	875
Quality	122 (1.4)	292 (1.3)	926
Interventions	117 (1.4)	339 (1.5)	968
Management	110 (1.3)	317 (1.4)	897
Self-management	106 (1.3)	304 (1.3)	901
Health information	105 (1.2)	260 (1.1)	797
Digital divide	97 (1.2)	258 (1.1)	805
Readability	90 (1.1)	201 (0.9)	671
Risk	87 (1.0)	263 (1.1)	645
Patient education	86 (1.0)	227 (1.0)	631

[a]N=8434.

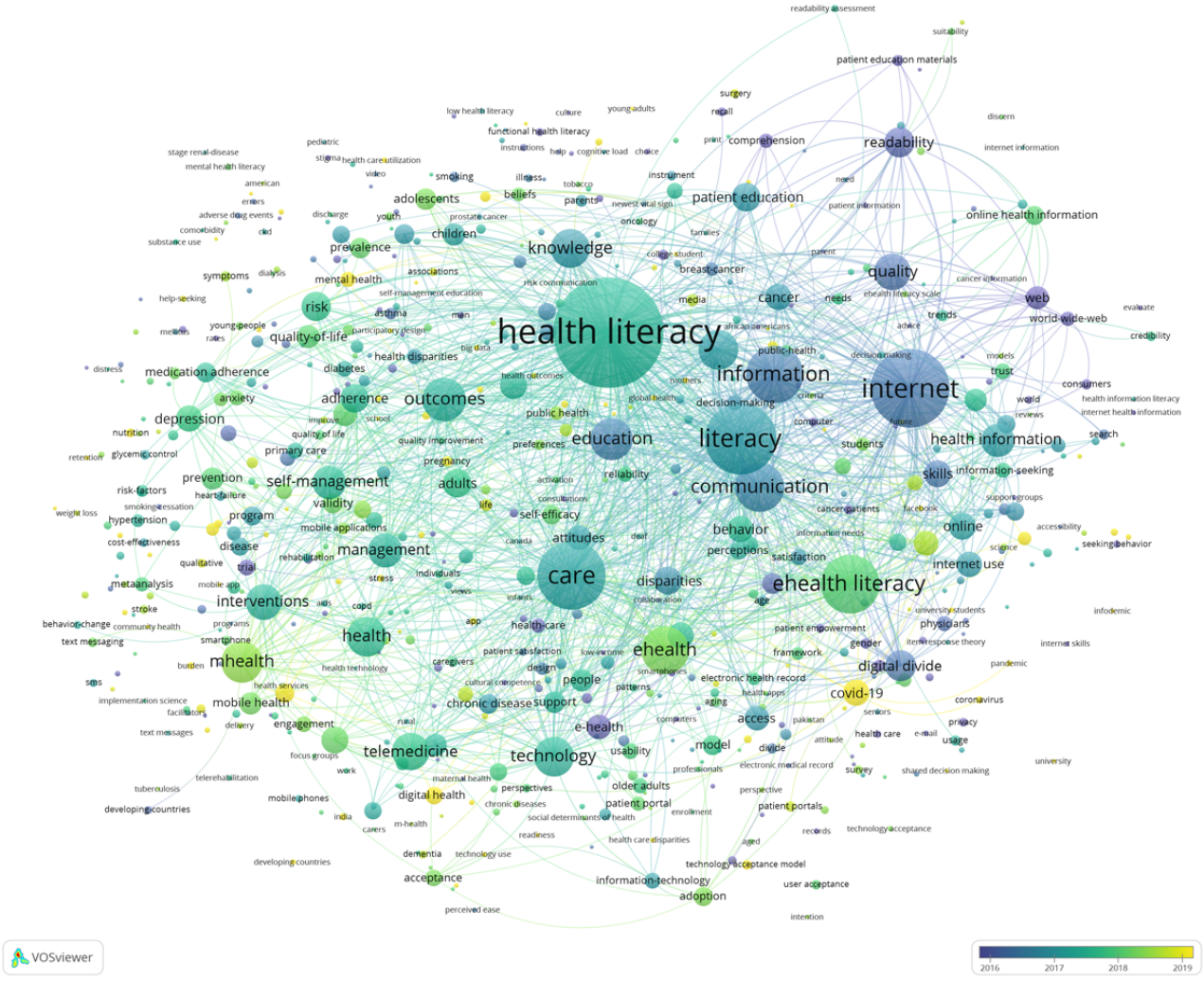
[b]N=23,124.

[c]Links refers to the number of keywords linked to a given keyword in the keyword co-occurrence network; total link strength refers to the total strength of the co-occurrence links of a given keyword with other keywords. The full list of co-occurrence keywords can be found in Multimedia Appendix 5.

As shown in Figure 7, the blue cluster mainly consists of health literacy keywords and the key factors affecting health literacy. The keywords include “health literacy” (n=666, 7.9%), “outcomes” (n=159, 1.9%), “education” (n=152, 1.8%), “knowledge” (n=134, 1.6%), and “self-management” (n=106, 1.3%). The core keywords that appear in the green cluster are “internet” (n=397, 4.7%), “literacy” (n=339, 4%), “eHealth literacy” (n=252, 3%), “information” (n=241, 2.9%), and “eHealth” (n=186, 2.2%). The co-occurrence intensity between these keywords is high, indicating that digital health literacy is closely related to internet and information literacy. The core keywords in the yellow cluster are “care” (n=316, 3.7%), “technology” (n=138, 1.6%), “telemedicine” (n=127, 1.5%), “digital divide” (n=97, 1.2%), “disparities” (n=81, 0.9%), “telehealth” (n=77, 0.9%), and “access” (n=67, 0.8%). These keywords form a co-occurrence network focused on the related technologies of care, the digital divide, disparities, and care access. The red cluster relates largely to digital health and has the core keywords “mHealth” (n=148, 1.8%), “health” (n=141, 1.7%), “interventions” (n=117, 1.4%), “risk” (n=87, 1%), and “depression” (n=74, 0.9%). These describe the detection of issues and health interventions as well as health risks using digital technology.

Figure 8 is the overlay visualization of keywords. In this figure, color represents average time of occurrence, node size represents occurrence frequencies, and link thickness represents co-occurrence strength (See Multimedia Appendix 7 for full-size image). Dark blue indicates an earlier appearance, and yellow indicates a more recent appearance. In terms of average time of occurrence, “information literacy” was the earliest keyword to appear, followed by “electronic health” and “mobile health,” which gradually changed to “digital health literacy” and expanded to include subfields like COVID-19 and mental health.

Figure 8. Overlay visualization of keyword co-occurrences.



Author Distribution

Table 7 presents the top 10 most productive authors. A total of 20 articles were published by Lyles, accounting for 1% of all 1955 articles. Lyles ranks first in terms of citations (n=386, 1.4%), with 19 links and a total link strength of 72, indicating that most articles were completed in collaboration with others. Wolf also published 20 (1%) articles. Schillinger published 17 articles (0.9%), ranking third for publications. Accordingly, we found that some of the most productive authors, namely, Wolf and Schillinger [25], were also the most productive in the field of health literacy. In addition, we investigated the h-index and found that Stellefson's research may have had great impaction in the fields of digital health literacy. The i10-index and g-index can be used to complement the h-index.

We also analyzed coauthorships using VOSviewer. The coauthorship network comprises 323 authors who have published more than 3 articles each (Figure 9). In the figure, color represents cluster, node size represents publications, and link thickness represents collaboration strength.

There are 103 clusters in total. The 4 most significant clusters are green, red, azure, and magenta. The green cluster includes Lyles (20 publications, total link strength 72), Schillinger (17 publications, total link strength 65), and Ratanawongsa (9 publications, total link strength 34). The high total link strength among collaborative authors indicates a strong co-operative relationship. All 3 authors are colleagues at the University of California San Francisco. The article with the highest number of citations analyzes the obstacles to and promoting factors for the use of online portals by patients and caregivers. The study found that participants with limited health literacy faced more basic barriers, including reading and typing challenges, limited personal experience with online security vulnerabilities/viruses, and distrust in potential security measures [37]. Another study they collaborated on suggested that the health literacy of a patient population was a powerful indicator of which patients needed the most support in using health technologies [38].

In the red cluster, Wolf (20 publications, total link strength 49) represents the core node, accompanied by O'Conor (7 publications, total link strength 23) and Federman (6 publications, total link strength 22). Their research focused on how health literacy affects access to and use of health technologies [39,40]. They also summarized clinical [41] and randomized clinical trials [42] to explore the impact of health literacy differences on clinical outcomes.

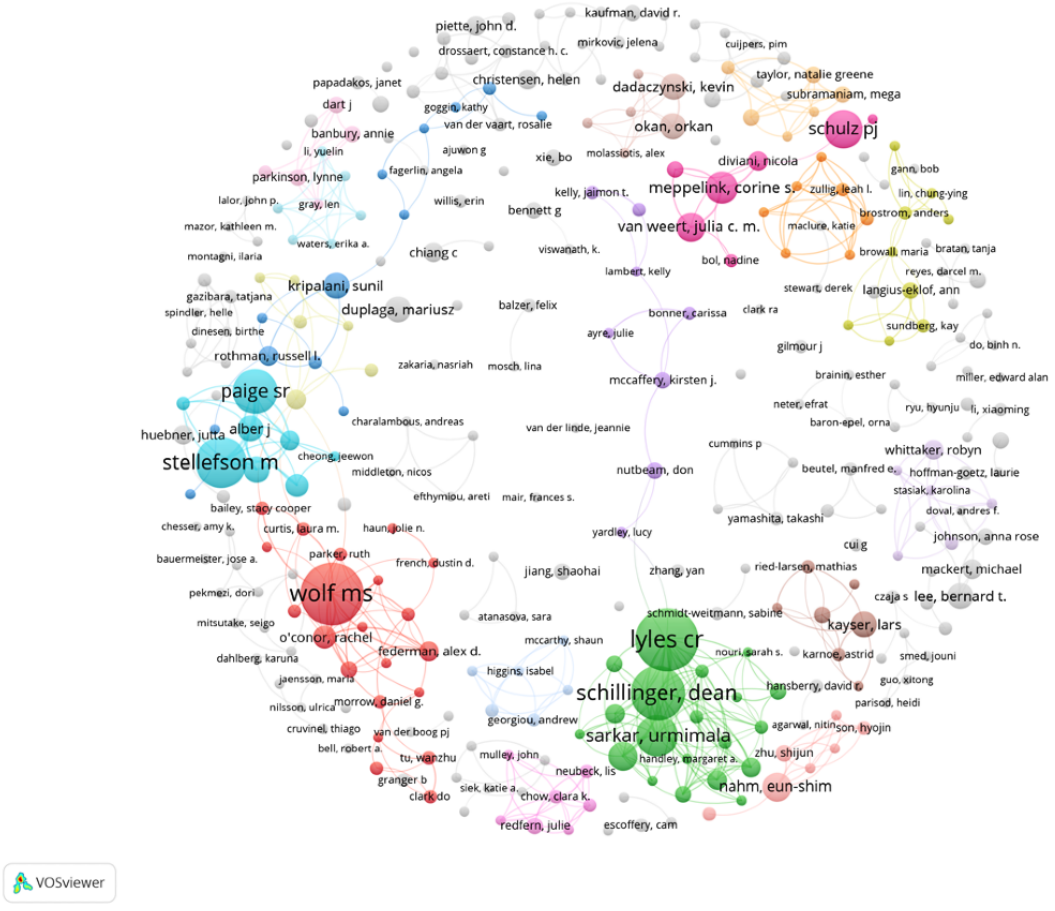
Table 7. Top 10 most productive authors.

Author	Publications[a], n (%)	Citations[b], n (%)	Links	Total link strength	i10-index	G-index	H-index
Lyles	20 (1)	386 (1.4)	19	72	7	19	8
Wolf	20 (1)	285 (1.1)	16	49	7	16	8
Schillinger	17 (0.9)	266 (1)	16	65	5	16	7
Stellefson	16 (0.8)	466 (1.7)	7	50	10	16	10
Paige	14 (0.7)	452 (1.7)	7	47	9	14	9
Sarkar	13 (0.7)	340 (1.3)	14	42	6	13	8
Schulz	12 (0.6)	224 (0.8)	2	4	8	12	8
Meppelink	10 (0.5)	253 (0.9)	4	17	6	10	7
Van Weert	9 (0.5)	265 (1)	4	14	6	9	7
Ratanawongsa	9 (0.5)	210 (0.8)	11	34	3	9	5

[a]N=1955.

[b]N=27,012.

Figure 9. Coauthorship network based on publications.



Stellefson (16 publications, total link strength 50) and Paige (14 publications, total link strength 47) form the core of the azure cluster. The co-operation intensity between them is high, and they are both from the University of Florida. Their research not only explores the factors that influence eHealth literacy and health information acquisition behavior [8,43] but also evaluates the reliability of eHEALS scores for patients with chronic illnesses [44]. They also examine literacy heterogeneity in the relationship between eHealth literacy and trust in online health communication channels and information sources [45] as well as test the measurement invariance of the eHEALS. Scholars believe that the eHEALS can be used to assess, monitor, and evaluate internet users' understanding of eHealth resources, information search skills, and participation abilities [46].

The magenta cluster includes Schulz (12 publications, total link strength 4), Meppelinks (10 publications, total link strength 17), and Van Weert (9 publications, total link strength 14). Schulz's co-operation strength is relatively low because most of his

collaborators published fewer than 3 articles. He tested the measurement invariance of the eHEALS across countries [47] and examined the reliability and validity of the Italian version of the eHEALS. Schulz's latest research analyzes the impact of subjective and objective health literacy on patients' judgment of health information and their decision-making ability [48].

Meppelinks and Van Weert, both from the University of Amsterdam, have a strong coauthorship relationship. Their articles analyze the impact of textual difficulty and illustrations on populations with low or high health literacy. They found that low health literacy groups benefit from well-illustrated information and nonchallenging texts, whereas high health literacy groups benefit from challenging texts [49]. They also analyzed the effectiveness of health animations among groups with differing health literacy. Narrated animations were revealed to be the best way to convey complex health information to people with low health literacy; this form can even bridge the information processing gap between low health literacy and high health literacy audiences [50]. Additionally, they explore the role of health literacy in the evaluation of online health information. Their study suggests that the differing evaluative abilities of people with different health literacies might be related to variances in issue perception and evaluation criteria [51]. They also evaluated the credibility, usefulness, and persuasiveness of both positive and negative texts on vaccination to examine the relationship between confirmation bias and health literacy in online health information searches. They determined that biased choices and biases against information persuasiveness were more prevalent among highly health-literate individuals, suggesting that bias recognition is important in the context of vaccination [52].

The number of articles published by an author does not reflect the quality or popularity of their work. We thus further analyzed the coauthorship network based on citations (Figure 10). Compared with Figure 9, each parameter in Figure 10 is the same except node size, which represents the number of citations. As shown in Figure 10, the 3 articles by Norman have been cited 995 times. His most notable contribution was the introduction of the concept of eHealth literacy in 2006. He defined eHealth literacy as the ability to seek, discover, understand, and evaluate health information from electronic sources and to apply the knowledge gained to solve health problems. As previously mentioned, his paper proposed the eHealth Literacy Scale, which includes 6 core literacies: traditional literacy, health literacy, information literacy, scientific literacy, media literacy, and computer literacy. This research was published in the *Journal of Medical Internet Research* and played a key role in the development of digital health literacy [8]. Another important paper by Norman detailed the design of the eHEALS, which not only evaluates consumers' perception of using information technology to promote health but also helps to determine whether eHealth programs match consumers' needs [12]. The scale is widely used in follow-up studies on digital health literacy.

In addition, Rothman and Russell coauthored 6 publications with a total of 449 citations. Their study in collaboration with DeWalt found that primary care–based heart failure self-management programs can reduce the risk of hospitalization or death among low health-literate patients [53]. Another study by Rothman and Russell suggested that many parents do not understand the common health information needed to take care of an infant. In light of this, a new Parental Health Literacy Activities Test (PHLAT) was developed to evaluate parents' health literacy [54]. In addition, Rothman and Russell also evaluated the reliability and effectiveness of Brief Health Literacy Screen (BHLS) in routine clinical settings [55].

To better reflect the impact and contributions of the authors' research, we conducted a co-citation analysis of the cited references in 1955 articles. The co-citation network was established by selecting authors with more than 10 citations. A total of 919 authors appeared in the network (selected from a total of 40,628), forming the 4 clusters shown in Figure 11. In this figure, color represents cluster, node size represents citations, and link thickness represents co-citation strength (see Multimedia Appendix 8 for full-size image).

Figure 11 shows that Norman represents the core node in the yellow cluster. Norman had the most citations and strongest co-citation strength. As previously mentioned, he introduced the electronic health literacy model [8] and designed the eHEALS [12] and thus played a very important role in the development of digital health literacy research. Another notable author is Nutbeam, who conducted various research on many aspects of digital health literacy, such as eHealth using with low levels of health literacy [56,57], attitudes of people with different health literacy toward digital health interventions, and skills for telehealth [58].

The WHO is the key node in the red cluster. As an agency author, the WHO has always played a pivotal role in the health practice field, including digital health literacy.

Fox is the most prominent author in the blue cluster. She worked on the Pew Research Center's internet project from 2000 to 2014 and mainly studied the impact of internet technology on health. During her time at the Pew Research Center, she published several highly cited research reports on health information [59], mobile health [60], health online [61], and other topics. From 2015 to 2017, she was the chief technology officer of the US Department of Health and Human Services. Another notable author in the blue cluster is Eysenbach, who made significant contributions to the definition of eHealth [62] and the evaluation of internet health information quality [63,64].

Baker, the most significant author in the green cluster, developed a simple scale measuring functional health literacy early in 1999 [65]. In 2006, he designed a concept model of health literacy that received widespread attention and played an important role in promoting the definition and measurement of health literacy [66]. Berkman is also an important author in the green cluster. He contributed to the definition of health literacy [67] and proposed that low health literacy is related to poor health outcomes and poor use of health care services [20]. The Centers for Disease Control and Prevention (CDC) are also in the green cluster, indicating that this agency has paid great attention to digital health literacy.

Figure 10. Coauthorship network of authors based on citations.

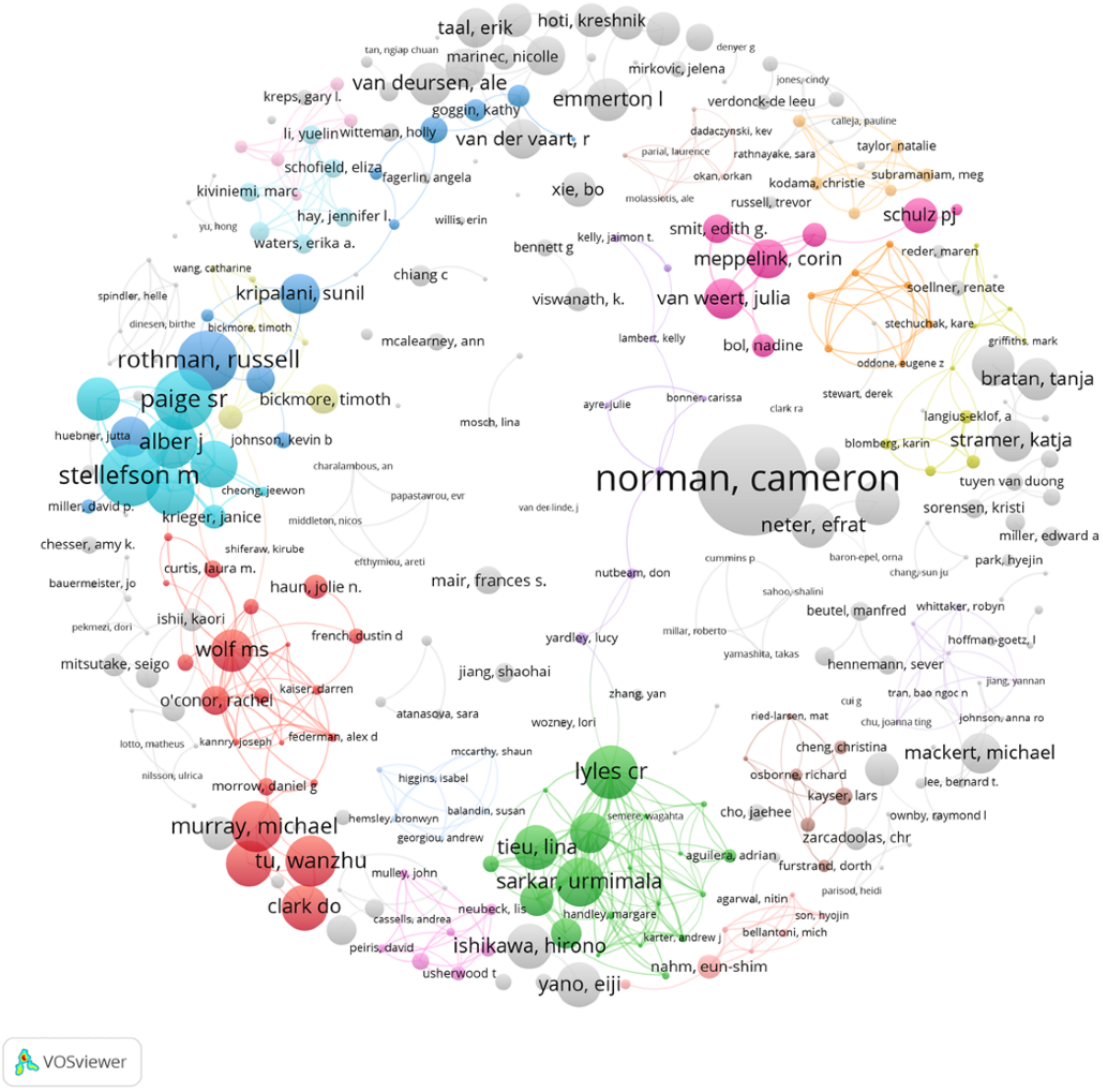
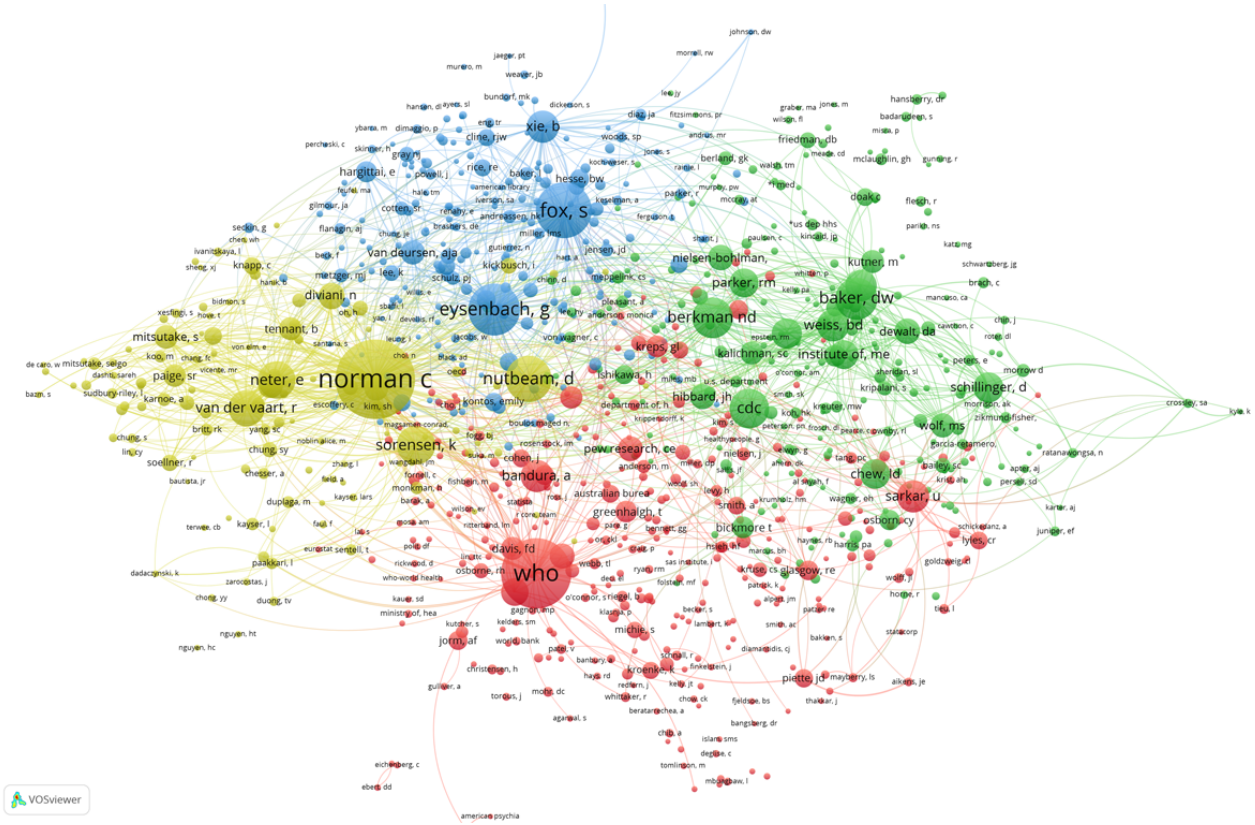


Figure 11. Co-citation network.



Articles With High Citation Numbers

Based on the keyword co-occurrence and author co-operation analysis, we further explored the cited articles. Table 8 reports the top 10 articles according to number of citations. We concluded that these 10 articles focus on 4 research themes: definition and measurement of digital health literacy, digital health literacy and health outcomes, digital health literacy and the digital divide, and the influencing factors of digital health literacy.

Two of Norman's articles are the most prominent with regard to definition and measurement of digital health literacy. As noted, the eHealth literacy conceptual model [8] and eHEALS [12] have contributed significantly to the study of eHealth literacy. Ishikawa et al [71] proposed and examined a newly developed scale that measures 3 different aspects of health literacy (ie, functional, communicative, and critical). This scale is a reliable and effective measurement method for the 3 types of health literacy in diabetic patients. Exploring patients' health literacy may lead to a better understanding of potential barriers to patients' self-management and health promoting behaviors [71].

Additionally, 3 articles discussed the theme of digital health literacy and health outcomes. Baker found that insufficient functional health literacy increases hospitalization risk [68]. Murray conducted a randomized clinical trial from a pharmacist intervention perspective and followed the trial with an electronic monitor, finding that patient health literacy significantly influenced the impact of pharmacist interventions on health outcomes [69]. DeWalt et al [53] determined that a primary care–based heart failure self-management program designed for low-literacy patients reduced risk of hospitalization or death.

Moreover, 2 articles analyzed the theme of digital health literacy and the digital divide. Neter and Brainin [22] explored the digital divide in relation to health information and argue that the internet reinforces existing social disparities. They maintain that a more comprehensive and sophisticated use of the internet alongside the growth of a highly eHealth–literate population has created new inequalities in the digital health information arena. Focusing on the digital divide among low-income home-based seniors, Choi and DiNitto surveyed internet use patterns, reasons for stopping use, eHealth literacy, and attitudes toward computer/internet use among low-income home-based seniors and younger adults under the age of 60. The study found that internet usage among this population is very low compared to the overall US population base due to either lack of access to computers and internet technologies, lack of financial resources to obtain computers and access technologies for personal use, or limitations due to medical conditions and disabilities [70].

Furthermore, 2 articles studied the influencing factors for digital health literacy. As mentioned earlier, Tennant et al [43] explored the extent to which sociodemographics, social determinants, and electronic device use influence eHealth literacy and the use of Web 2.0 for health information among baby boomers and older adults. Their study found that higher eHealth literacy among baby boomers and the older adults was related to being younger and more educated. Using a HealthSpace case study, Greenhalgh et al [72] found that policymakers hoped for patient empowerment, personalized care, lower health care costs, better data quality, and improved health literacy through personal

electronic health records. However, these records are not being used adequately. Because personal electronic health records must be closely aligned with people's attitudes, self-management practices, identified information needs, and health care options, the risk of them being abandoned or not adopted at all is significant.

Table 8. Top 10 articles according to number of citations.

Title	Author	Journal	Year	Citations	Reference	Theme
"eHealth literacy: Essential skills for consumer health in a networked world"	Norman and Skinner	*Journal of Medical Internet Research*	2006	722	[8]	Definition and measurement of digital health literacy
"Health literacy and the risk of hospital admission"	Baker et al	*Journal of General Internal Medicine*	1998	501	[68]	Digital health literacy and health outcomes
"Pharmacist intervention to improve medication adherence in heart failure - A randomized trial"	Murray et al	*Annals of Internal Medicine*	2007	297	[69]	Digital health literacy and health outcomes
"eHealth literacy: extending the digital divide to the realm of health information"	Neter and Brainin	*Journal of Medical Internet Research*	2012	282	[22]	Digital health literacy and the digital divide
"eHEALS: The eHealth Literacy Scale"	Norman and Skinner	*Journal of Medical Internet Research*	2006	271	[12]	Definition and measurement of digital health literacy
"eHealth Literacy and web 2.0 health information seeking behaviors among baby boomers and older adults"	Tennant et al	*Journal of Medical Internet Research*	2015	253	[43]	Influencing factors for digital health literacy
"The digital divide among low-income homebound older adults: internet use patterns, eHealth literacy, and attitudes toward computer/internet use"	Choi and DiNitto	*Journal of Medical Internet Research*	2013	248	[70]	Digital health literacy and the digital divide
"Measuring functional, communicative, and critical health literacy among diabetic patients"	Ishikawa et al	*Diabetes Care*	2008	220	[71]	Definition and measurement of digital health literacy
"A heart failure self-management program for patients of all literacy levels: A randomized, controlled trial" [ISRCTN11535170]	DeWalt et al	*BMC Health Services Research*	2006	213	[53]	Digital health literacy and health outcomes
"Adoption, non-adoption, and abandonment of a personal electronic health record: case study of HealthSpace"	Greenhalgh et al	*BMJ-British Medical Journal*	2010	174	[72]	Influencing factors for digital health literacy

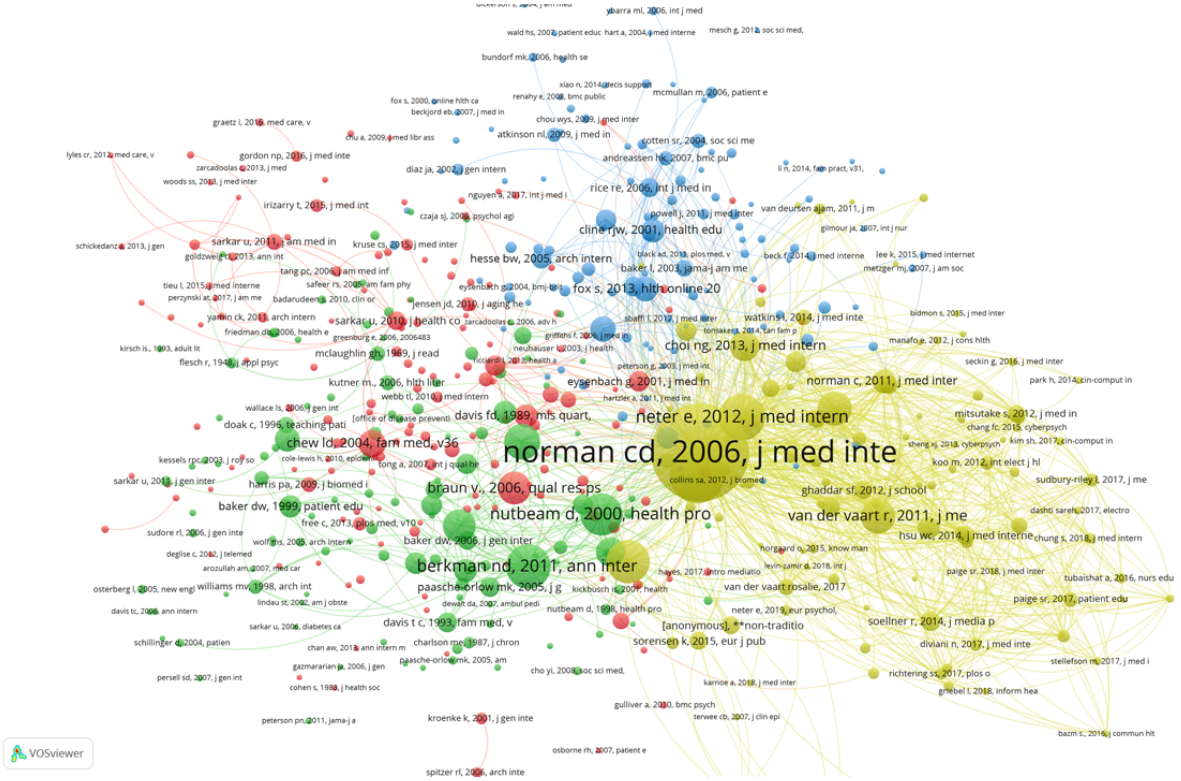
Co-cited Literature

To further examine literature that has played an important role in promoting the field of digital health literacy, we also conducted co-citation network analysis of references cited by the 1955 articles. A total of 510 references with more than 10 citations frequencies were selected as the co-citation network nodes, and 4 clusters were formed, as shown in Figure 12. In this figure, color represents cluster, node size represents citations, and link thickness represents co-citation strength. The full-size figure can be found in Multimedia Appendix 9.

The articles in the yellow cluster focus on the definition and scale of digital health literacy. Norman and Skinner [8,12] are at the core of the whole co-citation network. Within the yellow cluster, the article by Neter and Brainin [22] extends the digital divide to health information and health inequalities. Sørensen's [73] article promotes the development of definitions and conceptual models for health literacy. His research proposes a model integrating public views of medicine and health literacy. The model can be used as a conceptual basis for the development of interventions to improve health literacy and the development and validation of health literacy measurement tools. In addition, the paper tests by van der Vaart, the eHEALS by Norman, and the Dutch version of the eHEALS can be used to test for sufficient internal consistency and predictive validity [74].

In Figure 12, articles shown in the green cluster have supported research on health literacy and health outcomes. Berkman [20] proposed that low health literacy is related to poor health outcomes and poor use of health care services. Nutbeam [75] proposed a health outcomes model emphasizing health literacy an important result of health education. Nutbeam's study of health literacy as a concept showed the differences between functional, interactive, and critical health literacy and determined that improving health literacy entails not only transmitting information but also developing skills such as reading booklets and making successful appointments.

Figure 12. Co-citation network of cited references.



The red cluster includes articles that establish the theoretical basis and research methods of digital health literacy research. For example, Braun [76] advocates for thematic analysis as a useful and flexible method for qualitative psychological research. The testing scale of perceived usefulness and perceived ease of use developed by Davis [77] is very important and provides basic support for research on user acceptance. Digital health literacy often involves issues such as measuring users' acceptance of digital technology, and Davis' study has therefore been widely cited. A study by Chew [78] discussed the simple problem of identifying patients with poor health literacy and presented the Short Test of Functional Health Literacy in Adults (STOHFLA) scale, which has 3 key screening questions: "How often can someone help you read hospital information?" "How confident are you in filling out your own medical forms?" and "How often do you have questions about your health due to difficulty understanding written information?" These 3 questions effectively screen for health literacy inequality [78]. Hence, Chew's study provided a basis for later measurement of and scale construction for digital health literacy.

Research in the blue cluster focuses on health literacy and the digital divide. Cline [79] carried out a literature review and explored consumers' search for health information on the internet. Her paper discussed the criteria for evaluating online health information and argues that attention should be paid not only to "network gap" and information quality but also to the communication and transaction quality inherent in internet use [79]. Kontos [80] examined the use of eHealth tools according to sociodemographic factors such as race/ethnicity, socioeconomic status, age, and gender and found that there is a digital divide—compared with their peers, adults with poor economic conditions, adults who are older, and males are less likely to participate in a large number of eHealth activities [80].

The cited references include strong citation bursts reflecting changes in trending topics over time. A citation burst indicate the citations of an article increase rapidly within a period time. In this study, CiteSpace was used to analyze bursts of co-cited articles. The co-citation network was constructed based on the total number of the top 20 cited references each citation year. There were 1830 nodes and 6236 links. The 59 burst items found by CiteSpace are reported in Multimedia Appendix 10. Figure 13 reports the top 25 articles of 59 burst references with the strongest burstiness. The references are ranked by the starting year of the burst. In Figure 13, "references" in the first column refers to the list of publications with high burst; "year" in the second column refers to the year in which a given publication was first cited; "strength" refers to the citation burst strength of a given publication, which was calculated by the default Kleinberg algorithm of CiteSpace; "begin" refers to the year in which the citation of a publication begins to burst; and "end" refers to the end year of a bursting publication.

The article that burst earliest is that of Norman and Skinner [8]. In 2021, 2 articles burst, namely, those of van der Vaart et al [74] and Xie [81]. Van der Vaart et al [74] examined the validity of the Dutch version of the eHEALS, while Xie [81] tested the effect of electronic health literacy intervention for the elderly. Regardless of the specific learning method employed, the tested electronic health literacy intervention significantly improved the efficacy of digital health literacy and led to positive changes in self-managed health care [81]. Neter and Brainin burst suddenly from 2013 to 2017 and had a strong burstiness of 24.75 [22]. The burstiness of Tennant et al [43] was the highest among all cited articles. Their citations significantly burst from 2016 until 2021. In addition, the article published by Diviani in 2015 [82] burst at the same stage from 2016 to 2021.

In 2017, 5 articles were frequently cited. In 1 of these, Mitsutake [83] studied the relationship between eHealth literacy and the healthy behaviors of adult internet users. The study found that some healthy behaviors, including exercise and balanced nutrition, are independently related to eHealth literacy in Japan [83]. Citations of this article in 2017 burst and lasted until 2021, with a high bursting strength of 20.39.

In 2018, 4 articles had the strongest citation bursts, which lasted until 2021. Paige [44] assessed the reliability of the eHEALS for patients with chronic diseases. Van der Vaart [6] argued that previous tools for measuring digital health literacy focused on information collection (Health 1.0 skills) but not on network interactions (Health 2.0) and therefore developed the new Digital Health Literacy Instrument (DHLI) to measure operating skills, navigation skills, information searching, evaluation of reliability, determining relevance, adding self-generated content, and protecting privacy [6]. The burst duration of this article lasted from 2018 to 2021. Diviani [13] examined the reliability and validity of the Italian version of the eHEALS (I-eHEALS). Perez [14] validated the Spanish version of the eHEALS.

Of the 4 articles that burst in 2019, 1 was a literature review, while the other 3 were extensions of the eHEALS study. Kim [84] conducted a literature review on health literacy in the electronic age. Sudbury-Riley [47] tested the multinational test invariance of the eHEALS; Chung [16] developed the Korean version of the eHEALS; and Tubaishat [85] studied electronic health literacy among nursing students.

Figure 13. Top 25 references with the strongest citation bursts among a co-citation network of cited references.

Discussion

Principal Findings

In this study, we conducted a comprehensive bibliometric analysis of the field of digital health literacy. A total of 1955 scientific publications were retrieved from the WoS Core Collection, and the development of digital health literacy over the past 20 years was analyzed. This paper systematically summarizes the development context and trends of digital health literacy research, and our work serves as a basic reference and directional guide for future research in this field. Our study uncovered the most productive countries, institutions, journals, and authors. It also identified the different stages of research trends in digital health literacy, the most important research themes of digital health literacy, and the evolution of the concept of digital health literacy.

Our analysis of national and regional contributions revealed that the United States is the leader in this field in terms of both publications and citations. In relation to institutional contributions, the University of California System is in the top position. Our analysis of institutional collaboration networks shows that the focus on digital health literacy research is gradually expanding from high-income countries to emerging market countries and regions. The *Journal of Medical Internet Research* is the leading journal in digital health literacy research for number of articles, citations, and co-citations. In terms of author contributions, Lyles, Wolf, and Schillinger are the top 3 authors, while Norman has the highest number of citations. Among them, Wolf and Schillinger are also the most productive in the area of health literacy [25]. Regarding the co-citation of authors, Norman is the most co-cited individual author, and the WHO is the most co-cited institutional author.

Research on digital health literacy spans from 1998 to 2021. In terms of annual publications, the research history can be divided into 3 phases: (1) 1998 to 2005 was the incubation period, with almost no growth in the number of publications and the number of publications remaining in the single digits; (2) 2006 to 2013

was the slow growth period, with the number of publications remaining below 100; (3) after 2014, digital health literacy research entered a period of rapid growth, with the annual number of articles exceeding 100.

Thematic analysis demonstrated that health care sciences services, medical informatics, and public environmental occupational health are the 3 most important research fields related to digital health literacy. The theme bursting analysis revealed that early studies mainly focused on electronic medical records (widely cited from 2011 to 2019) and electronic health records (widely cited from 2013 to 2018), indicating that studies began to focus on topics such as health records and health information before 2014; this was also the gestation period for the concept of digital health literacy. The second bursting peak occurred in 2014 and has continued into the present. The theme terms related to health literacy and health information in this stage have become trending topics. In 2015, the eHEALS began to burst. Subsequently, thematic terms such as eHealth literacy began to emerge and trend. After 2014, the concept of digital health literacy gradually matured, and relevant studies began to explode. From the evolution of the theme's citation bursts, concepts related to digital health literacy evolved from electronic medical record, electronic health record, eHealth Literacy Scale, to eHealth literacy. This evolution can be attributed to the advancement of digital technology and the digital health industry. In the early days, the use of electronic technology only turned paper health records into electronic health information, such as electronic medical records and electronic health records. With the escalating empowerment of the health industry by digital technology, especially the emergence of professional measurement tools such as the eHealth Literacy Scale, a moniker related to digital health literacy has gradually emerged. The emergence and evolution of this kind of terminology are in line with the general pattern of an emerging field of study from initiation to maturity.

The co-occurrence analysis of keywords yielded 4 clusters: keywords of health literacy and critical factors affecting health literacy; keywords of digital health literacy related to the internet and information literacy; keywords of related technologies to digital health literacy, the digital divide, disparities, and availability of care; and keywords related to the application of digital health. Based on the average frequencies of occurrence of keywords, information literacy was the earliest, followed by electronic health, mobile health, and digital health literacy, which expanded to digital health literacy niche research areas such as COVID-19 and mental health.

Our analysis of articles with high citations showed that they mainly focus on 4 aspects of digital health literacy: the definition and scale of digital health literacy [12], digital health literacy and health outcomes [53], digital health literacy and the digital divide [22,23], and influencing factors of digital health literacy [43]. Reference co-citation analysis revealed 4 clusters as well as the theoretical basis and research methods of digital health literacy research [77].

Comparisons With Prior Work

Some scholars have conducted bibliometric analysis on health literacy and eHealth literacy. Among them, a few suggest that mental health literacy and eHealth literacy will be 2 expansion directions for future health literacy research [25]. In contrast to previous bibliometric articles in related fields, our study is a comprehensive bibliometric analysis of the field of digital health literacy, which includes studies on eHealth literacy. There are also early publications that conducted bibliometric analyses on internet health information–seeking behaviors [86] but did not further explore the relationship between this behavior and health literacy. Some scholars also conducted a bibliometric analysis of consumer health informatics (CHIN) and found that research topics focused on patient education, health information demands, health information search behavior, health behavior interventions, health literacy, health information technology, and eHealth [87]. However, health informatics is broader in scope, and our study focuses more on the emerging subfield of digital health literacy. In addition, a recent bibliometric analysis of global eHealth found that one of the frontier issues in global eHealth research is the eHealth Literacy Scale [26]. This coincides with the research on digital health literacy measurement tools summarized in this paper. Some scholars likewise used bibliometric methods to analyze research hotspots and trends in eHealth literacy, arguing that eHealth literacy research faces challenges such as the development of terminological connotations, the objectivity of assessment methodology, and the impact of interventions [27]. However, we extended eHealth literacy to a broader scope: digital health literacy. In addition to encompassing eHealth literacy, we further expanded the scope of the literature retrieval by using the intersection of digital technology and health literacy and the intersection of digital health and literacy. Therefore, this study provides a systematic analysis for the development of research on digital health literacy.

Limitations

There are 3 main limitations of this study. First, in terms of data selection, the WoS core collection was selected, while other databases such as Scopus were not included. This study mainly focused on journals, and less attention was paid to other means of scientific knowledge dissemination (such as books, working papers, and reports). Therefore, some important studies may have been missed, especially emerging research.

Second, there are some subjective aspects in the sample selection in this paper. For example, on the one hand, we only analyzed English publications; thus, there may have been some linguistic bias. Future comparisons for articles published in different language or countries can be made. On the other hand, subjectivity may have influenced our search strategies and screenings as this is difficult to avoid in bibliometric studies.

Third, our study focused on a bibliometric analysis aimed at analyzing the structure of knowledge in the field of digital health literacy. There was no detailed discussion of study content. This calls for a more systematic literature review in the future.

Future Directions

The aforementioned results and discussion reveal that there are many issues in the field of digital health literacy that deserve further study. In particular, considering the ongoing COVID-19 pandemic, digital health literacy deserves more attention from

society and scholars. For example, it is important to consider how the government can promote health outcomes by enhancing the digital health literacy of the public to enhance COVID-19 prevention and control; how the digital health literacy of different groups can be measured to achieve effective COVID-19 prevention and control; and how the digital divide can be bridged by enhancing the digital health literacy of vulnerable groups to mitigate any health inequities caused by COVID-19.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Full-size Figure 3.
[PDF File (Adobe PDF File), 644 KB - jmir_v24i7e35816_app1.pdf]

Multimedia Appendix 2
Full-size Figure 4.
[PDF File (Adobe PDF File), 658 KB - jmir_v24i7e35816_app2.pdf]

Multimedia Appendix 3
Full-size Figure 5.
[PDF File (Adobe PDF File), 814 KB - jmir_v24i7e35816_app3.pdf]

Multimedia Appendix 4
Theme terms citation burst for Figure 6.
[PDF File (Adobe PDF File), 205 KB - jmir_v24i7e35816_app4.pdf]

Multimedia Appendix 5
Top 100 co-occurrence of keywords.
[XLS File (Microsoft Excel File), 43 KB - jmir_v24i7e35816_app5.xls]

Multimedia Appendix 6
Full-size Figure 7.
[PDF File (Adobe PDF File), 657 KB - jmir_v24i7e35816_app6.pdf]

Multimedia Appendix 7
Full-size Figure 8.
[PDF File (Adobe PDF File), 640 KB - jmir_v24i7e35816_app7.pdf]

Multimedia Appendix 8
Full-size Figure 11.
[PDF File (Adobe PDF File), 688 KB - jmir_v24i7e35816_app8.pdf]

Multimedia Appendix 9
Full-size Figure 12.
[PDF File (Adobe PDF File), 753 KB - jmir_v24i7e35816_app9.pdf]

Multimedia Appendix 10
References citation burst for Figure 13.
[PDF File (Adobe PDF File), 833 KB - jmir_v24i7e35816_app10.pdf]

References

1. Global strategy on digital health 2020-2025. World Health Organization. Geneva, Switzerland; 2020. URL: https://apps.who.int/iris/handle/10665/344249 [accessed 2022-06-26]
2. Shuren J, Patel B, Gottlieb S. FDA regulation of mobile medical apps. JAMA 2018 Jul 24;320(4):337-338. [doi: 10.1001/jama.2018.8832] [Medline: 29971339]

3. Kickbusch I, Piselli D, Agrawal A, Balicer R, Banner O, Adelhardt M, et al. The Lancet and Financial Times Commission on governing health futures 2030: growing up in a digital world. Lancet 2021 Nov;398(10312):1727-1776. [doi: 10.1016/s0140-6736(21)01824-9]
4. van Kessel R, Wong BLH, Clemens T, Brand H. Digital health literacy as a super determinant of health: More than simply the sum of its parts. Internet Interv 2022 Mar;27:100500 [FREE Full text] [doi: 10.1016/j.invent.2022.100500] [Medline: 35242586]
5. Nutbeam D. Health promotion glossary. Health Promot Int 1998 Jan 01;13(4):349-364. [doi: 10.1093/heapro/13.4.349]
6. van der Vaart R, Drossaert C. Digital Health Literacy Instrument. J Med Internet Res 2017 Jan 24;19(1):e27 [FREE Full text] [doi: 10.2196/jmir.6709] [Medline: 28119275]
7. European Citizens' Digital Health Literacy.: European Commission; 2015. URL: https://data.europa.eu/doi/10.2759/86596 [accessed 2022-06-26]
8. 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]
9. Adil A, Usman A, Jalil A. Qualitative analysis of digital health literacy among university students in Pakistan. J Hum Behav Soc Environ 2020 Dec 08;31(6):771-781. [doi: 10.1080/10911359.2020.1812462]
10. van der Vaart R, van Driel D, Pronk K, Paulussen S, Te Boekhorst S, Rosmalen JGM, et al. The role of age, education, and digital health literacy in the usability of internet-based cognitive behavioral therapy for chronic pain: mixed methods study. JMIR Form Res 2019 Nov 21;3(4):e12883 [FREE Full text] [doi: 10.2196/12883] [Medline: 31750839]
11. Patil U, Kostareva U, Hadley M, Manganello JA, Okan O, Dadaczynski K, et al. Health literacy, digital health literacy, and COVID-19 pandemic attitudes and behaviors in U.S. college students: implications for interventions. Int J Environ Res Public Health 2021 Mar 23;18(6):3301 [FREE Full text] [doi: 10.3390/ijerph18063301] [Medline: 33806763]
12. 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]
13. Diviani N, Dima AL, Schulz PJ. A psychometric analysis of the Italian version of the eHealth Literacy Scale using item response and classical test theory methods. J Med Internet Res 2017 Apr 11;19(4):e114 [FREE Full text] [doi: 10.2196/jmir.6749] [Medline: 28400356]
14. Paramio PG, Almagro BJ, Hernando, Aguaded GJI. Validation of the eHealth Literacy Scale (eHEALS) in Spanish university students. Rev Esp Salud Publica 2015 Jun;89(3):329-338 [FREE Full text] [doi: 10.4321/S1135-57272015000300010] [Medline: 26388346]
15. Soellner R, Huber S, Reder M. The concept of eHealth literacy and its measurement. J Media Psychol 2014 Jan;26(1):29-38. [doi: 10.1027/1864-1105/a000104]
16. Chung S, Park BK, Nahm E. The Korean eHealth Literacy Scale (K-eHEALS): reliability and validity testing in younger adults recruited online. J Med Internet Res 2018 Apr 20;20(4):e138. [doi: 10.2196/jmir.8759] [Medline: 29678800]
17. 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]
18. Karnoe A, Furstrand D, Christensen KB, Norgaard O, Kayser L. Assessing competencies needed to engage with digital health services: development of the eHealth literacy assessment toolkit. J Med Internet Res 2018 May 10;20(5):e178 [FREE Full text] [doi: 10.2196/jmir.8347] [Medline: 29748163]
19. Nutbeam D, Lloyd JE. Understanding and responding to health literacy as a social determinant of health. Annu Rev Public Health 2021 Apr 01;42(1):159-173. [doi: 10.1146/annurev-publhealth-090419-102529] [Medline: 33035427]
20. 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]
21. Miller DP, Denizard-Thompson N, Weaver KE, Case LD, Troyer JL, Spangler JG, et al. Effect of a digital health intervention on receipt of colorectal cancer screening in vulnerable patients. Ann Intern Med 2018 Mar 13;168(8):550. [doi: 10.7326/m17-2315]
22. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
23. Knapp C, Madden V, Wang H, Sloyer P, Shenkman E. Internet use and eHealth literacy of low-income parents whose children have special health care needs. J Med Internet Res 2011 Sep;13(3):e75 [FREE Full text] [doi: 10.2196/jmir.1697] [Medline: 21960017]
24. Dadaczynski K, Okan O, Messer M, Leung AYM, Rosário R, Darlington E, et al. Digital health literacy and web-based information-seeking behaviors of university students in Germany during the COVID-19 pandemic: cross-sectional survey study. J Med Internet Res 2021 Jan 15;23(1):e24097 [FREE Full text] [doi: 10.2196/24097] [Medline: 33395396]
25. Qi S, Hua F, Xu S, Zhou Z, Liu F. Trends of global health literacy research (1995-2020): Analysis of mapping knowledge domains based on citation data mining. PLoS One 2021 Aug 9;16(8):e0254988 [FREE Full text] [doi: 10.1371/journal.pone.0254988] [Medline: 34370749]

26. Tian H, Chen J. A bibliometric analysis on global eHealth. Digit Health 2022 Apr 11;8:20552076221091352 [FREE Full text] [doi: 10.1177/20552076221091352] [Medline: 35433016]
27. Wang C, Wu X, Qi H. A comprehensive analysis of e-health literacy research focuses and trends. Healthcare (Basel) 2021 Dec 30;10(1) [FREE Full text] [doi: 10.3390/healthcare10010066] [Medline: 35052230]
28. Chen C. Science mapping: a systematic review of the literature. J Data Inf Sci 2017 Mar 21;2(2):1-40. [doi: 10.1515/jdis-2017-0006]
29. Pritchard A. Statistical bibliography or bibliometrics? J Doc 1969;25(4):348-349. [doi: 10.1017/upo9788175968608.013]
30. 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. [doi: 10.2196/18228]
31. Ahmadvand A, Kavanagh D, Clark M, Drennan J, Nissen L. Trends and visibility of "digital health" as a keyword in articles by JMIR Publications in the new millennium: bibliographic-bibliometric analysis. J Med Internet Res 2019 Dec 19;21(12):e10477 [FREE Full text] [doi: 10.2196/10477] [Medline: 31855190]
32. Yang K, Qi H. Research on health disparities related to the COVID-19 pandemic: a bibliometric analysis. Int J Environ Res Public Health 2022 Jan 22;19(3):1220 [FREE Full text] [doi: 10.3390/ijerph19031220] [Medline: 35162243]
33. 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]
34. Chen C. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. J Am Soc Inf Sci 2006 Feb 01;57(3):359-377. [doi: 10.1002/asi.20317]
35. Garfield E. From the science of science to Scientometrics visualizing the history of science with HistCite software. J Informetr 2009 Jul;3(3):173-179. [doi: 10.1016/j.joi.2009.03.009]
36. Li T, Ho Y, Li C. Bibliometric analysis on global Parkinson's disease research trends during 1991-2006. Neurosci Lett 2008 Aug 29;441(3):248-252. [doi: 10.1016/j.neulet.2008.06.044] [Medline: 18582532]
37. 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]
38. Tieu L, Schillinger D, Sarkar U, Hoskote M, Hahn KJ, Ratanawongsa N, et al. Online patient websites for electronic health record access among vulnerable populations: portals to nowhere? J Am Med Inform Assoc 2017 Apr 01;24(e1):e47-e54. [doi: 10.1093/jamia/ocw098] [Medline: 27402138]
39. Bailey SC, O'Conor R, Bojarski EA, Mullen R, Patzer RE, Vicencio D, et al. Literacy disparities in patient access and health-related use of Internet and mobile technologies. Health Expect 2015 Dec;18(6):3079-3087 [FREE Full text] [doi: 10.1111/hex.12294] [Medline: 25363660]
40. 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]
41. Salmon C, O'Conor R, Singh S, Ramaswamy R, Kannry J, Wolf MS, et al. Characteristics of outpatient clinical summaries in the United States. Int J Med Inform 2016 Oct;94:75-80. [doi: 10.1016/j.ijmedinf.2016.06.005] [Medline: 27573314]
42. Federman AD, O'Conor R, Mindlis I, Hoy-Rosas J, Hauser D, Lurio J, et al. Effect of a self-management support intervention on asthma outcomes in older adults: the SAMBA study randomized clinical trial. JAMA Intern Med 2019 Aug 01;179(8):1113-1121 [FREE Full text] [doi: 10.1001/jamainternmed.2019.1201] [Medline: 31180474]
43. 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]
44. Paige SR, Krieger JL, Stellefson M, Alber JM. eHealth literacy in chronic disease patients: An item response theory analysis of the eHealth literacy scale (eHEALS). Patient Educ Couns 2017 Feb 16;100(2):320-326 [FREE Full text] [doi: 10.1016/j.pec.2016.09.008] [Medline: 27658660]
45. Paige SR, Krieger JL, Stellefson ML. The influence of eHealth literacy on perceived trust in online health communication channels and sources. J Health Commun 2017 Jan;22(1):53-65 [FREE Full text] [doi: 10.1080/10810730.2016.1250846] [Medline: 28001489]
46. Paige SR, Miller MD, Krieger JL, Stellefson M, Cheong J. Electronic health literacy across the lifespan: measurement invariance study. J Med Internet Res 2018 Jul 09;20(7):e10434 [FREE Full text] [doi: 10.2196/10434] [Medline: 29986848]
47. 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]
48. Schulz PJ, Pessina A, Hartung U, Petrocchi S. Effects of objective and subjective health literacy on patients' accurate judgment of health information and decision-making ability: survey study. J Med Internet Res 2021 Jan 21;23(1):e20457 [FREE Full text] [doi: 10.2196/20457] [Medline: 33475519]
49. Meppelink CS, Smit EG, Buurman BM, van Weert JCM. Should we be afraid of simple messages? the effects of text difficulty and illustrations in people with low or high health literacy. Health Commun 2015 Sep 15;30(12):1181-1189. [doi: 10.1080/10410236.2015.1037425] [Medline: 26372031]

50. Meppelink CS, Bol N. Exploring the role of health literacy on attention to and recall of text-illustrated health information: An eye-tracking study. Comput Hum Behav 2015 Jul;48:87-93. [doi: 10.1016/j.chb.2015.01.027]
51. Diviani N, van den Putte B, Meppelink CS, van Weert JC. Exploring the role of health literacy in the evaluation of online health information: Insights from a mixed-methods study. Patient Educ Couns 2016 Jun;99(6):1017-1025. [doi: 10.1016/j.pec.2016.01.007] [Medline: 26817407]
52. Meppelink CS, Smit EG, Fransen ML, Diviani N. "I was right about vaccination": confirmation bias and health literacy in online health information seeking. J Health Commun 2019 Mar 21;24(2):129-140. [doi: 10.1080/10810730.2019.1583701] [Medline: 30895889]
53. DeWalt DA, Malone RM, Bryant ME, Kosnar MC, Corr KE, Rothman RL, et al. A heart failure self-management program for patients of all literacy levels: a randomized, controlled trial [ISRCTN11535170]. BMC Health Serv Res 2006 Mar 13;6:30 [FREE Full text] [doi: 10.1186/1472-6963-6-30] [Medline: 16533388]
54. Kumar D, Sanders L, Perrin EM, Lokker N, Patterson B, Gunn V, et al. Parental understanding of infant health information: health literacy, numeracy, and the Parental Health Literacy Activities Test (PHLAT). Acad Pediatr 2010;10(5):309-316 [FREE Full text] [doi: 10.1016/j.acap.2010.06.007] [Medline: 20674532]
55. Willens DE, Kripalani S, Schildcrout JS, Cawthon C, Wallston K, Mion LC, et al. Association of brief health literacy screening and blood pressure in primary care. J Health Commun 2013;18 Suppl 1:129-142 [FREE Full text] [doi: 10.1080/10810730.2013.825663] [Medline: 24093351]
56. Parker SM, Stocks N, Nutbeam D, Thomas L, Denney-Wilson E, Zwar N, et al. Preventing chronic disease in patients with low health literacy using eHealth and teamwork in primary healthcare: protocol for a cluster randomised controlled trial. BMJ Open 2018 Jun 04;8(6):e023239 [FREE Full text] [doi: 10.1136/bmjopen-2018-023239] [Medline: 29866737]
57. Rowsell A, Muller I, Murray E, Little P, Byrne CD, Ganahl K, et al. Views of people with high and low levels of health literacy about a digital intervention to promote physical activity for diabetes: a qualitative study in five countries. J Med Internet Res 2015 Oct 12;17(10):e230 [FREE Full text] [doi: 10.2196/jmir.4999] [Medline: 26459743]
58. Morony S, Weir K, Duncan G, Biggs J, Nutbeam D, Mccaffery KJ. Enhancing communication skills for telehealth: development and implementation of a Teach-Back intervention for a national maternal and child health helpline in Australia. BMC Health Serv Res 2018 Mar 07;18(1):162 [FREE Full text] [doi: 10.1186/s12913-018-2956-6] [Medline: 29514642]
59. Fox S. The Social Life of Health Information.: Pew Research Center's Internet & American Life Project; 2011 May 12. URL: https://www.pewresearch.org/internet/2011/05/12/the-social-life-of-health-information-2011/ [accessed 2022-06-26]
60. Fox S, Duggan M. Mobile Health 2012.: Pew Research Center's Internet & American Life Project; 2012 Nov 8. URL: https://www.pewresearch.org/internet/2012/11/08/mobile-health-2012/ [accessed 2022-06-26]
61. Fox S, Duggan M. Health Online 2013.: Pew Research Center's Internet & American Life Project; 2013 Jan 15. URL: https://www.pewresearch.org/internet/2013/01/15/health-online-2013/ [accessed 2022-06-26]
62. 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]
63. Eysenbach G, Powell J, Kuss O, Sa E. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA 2002;287(20):2691-2700. [doi: 10.1001/jama.287.20.2691] [Medline: 12020305]
64. Eysenbach G, Köhler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. BMJ 2002 Mar 09;324(7337):573-577 [FREE Full text] [doi: 10.1136/bmj.324.7337.573] [Medline: 11884321]
65. Baker DW, Williams MV, Parker RM, Gazmararian JA, Nurss J. Development of a brief test to measure functional health literacy. Patient Educ Couns 1999 Sep;38(1):33-42. [doi: 10.1016/s0738-3991(98)00116-5]
66. Baker DW. The meaning and the measure of health literacy. J Gen Intern Med 2006 Aug;21(8):878-883 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00540.x] [Medline: 16881951]
67. Berkman ND, Davis TC, McCormack L. Health literacy: what is it? J Health Commun 2010 Aug;15 Suppl 2:9-19. [doi: 10.1080/10810730.2010.499985] [Medline: 20845189]
68. Baker DW, Parker RM, Williams MV, Clark WS. Health literacy and the risk of hospital admission. J Gen Intern Med 1998 Dec;13(12):791-798 [FREE Full text] [doi: 10.1046/j.1525-1497.1998.00242.x] [Medline: 9844076]
69. Murray MD, Young J, Hoke S, Tu W, Weiner M, Morrow D, et al. Pharmacist intervention to improve medication adherence in heart failure: a randomized trial. Ann Intern Med 2007 May 15;146(10):714-725. [doi: 10.7326/0003-4819-146-10-200705150-00005] [Medline: 17502632]
70. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: Internet use patterns, eHealth literacy, and attitudes toward computer/Internet use. J Med Internet Res 2013 May 02;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]
71. Ishikawa H, Takeuchi T, Yano E. Measuring functional, communicative, and critical health literacy among diabetic patients. Diabetes Care 2008 May 01;31(5):874-879. [doi: 10.2337/dc07-1932] [Medline: 18299446]
72. Greenhalgh T, Hinder S, Stramer K, Bratan T, Russell J. Adoption, non-adoption, and abandonment of a personal electronic health record: case study of HealthSpace. BMJ 2010 Nov 16;341(nov16 1):c5814-c5814 [FREE Full text] [doi: 10.1136/bmj.c5814] [Medline: 21081595]

73. Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, (HLS-EU) Consortium Health Literacy Project European. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012 Jan 25;12:80 [FREE Full text] [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]
74. van der Vaart R, van Deursen AJ, Drossaert CH, Taal E, van Dijk JA, van de Laar MA. Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations. J Med Internet Res 2011 Nov 09;13(4):e86 [FREE Full text] [doi: 10.2196/jmir.1840] [Medline: 22071338]
75. Nutbeam D. Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. Health Prom Int 2000;15(3):a. [doi: 10.1093/heapro/15.3.259]
76. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
77. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008]
78. Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. Fam Med 2004 Sep;36(8):588-594 [FREE Full text] [Medline: 15343421]
79. Cline R, Haynes K. 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]
80. Kontos E, Blake KD, Chou WS, 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]
81. Xie B. Effects of an eHealth literacy intervention for older adults. J Med Internet Res 2011 Nov 03;13(4):e90 [FREE Full text] [doi: 10.2196/jmir.1880] [Medline: 22052161]
82. Diviani N, van DPB, Giani S, van WJC. Low health literacy and evaluation of online health information: a systematic review of the literature. J Med Internet Res 2015 May;17(5):e112 [FREE Full text] [doi: 10.2196/jmir.4018] [Medline: 25953147]
83. Mitsutake S, Shibata A, Ishii K, Oka K. Associations of eHealth Literacy With Health Behavior Among Adult Internet Users. J Med Internet Res 2016 Jul 18;18(7):e192 [FREE Full text] [doi: 10.2196/jmir.5413] [Medline: 27432783]
84. Kim H, Xie B. Health literacy in the eHealth era: A systematic review of the literature. Patient Educ Couns 2017 Jun;100(6):1073-1082. [doi: 10.1016/j.pec.2017.01.015] [Medline: 28174067]
85. Tubaishat A, Habiballah L. eHealth literacy among undergraduate nursing students. Nurse Educ Today 2016 Jul;42:47-52. [doi: 10.1016/j.nedt.2016.04.003] [Medline: 27237352]
86. Li F, Li M, Guan P, Ma S, Cui L. Mapping publication trends and identifying hot spots of research on Internet health information seeking behavior: a quantitative and co-word biclustering analysis. J Med Internet Res 2015 Mar 25;17(3):e81 [FREE Full text] [doi: 10.2196/jmir.3326] [Medline: 25830358]
87. Ouyang W, Xie W, Xin Z, He H, Wen T, Peng X, et al. Evolutionary overview of consumer health informatics: bibliometric study on the Web of Science from 1999 to 2019. J Med Internet Res 2021 Sep 09;23(9):e21974 [FREE Full text] [doi: 10.2196/21974] [Medline: 34499042]

Abbreviations

A&HCI: Art and Humanities Citation Index
BHLS: Brief Health Literacy Screen
CDC: Centers for Disease Control and Prevention
CHIN: consumer health informatics
DHLI: Digital Health Literacy Instrument
eHEALS: eHealth Literacy Scale
FDA: Food and Drug Administration
mHealth: mobile health
PHLAT: Parental Health Literacy Activities Test
PLOS: Public Library of Science
SCIE: Science Citation Index Extension
SSCI: Social Science Citation Index
STOHFLA: Short Test of Functional Health Literacy in Adults
WHO: World Health Organization
WoS: Web of Science

Edited by T Leung; submitted 18.12.21; peer-reviewed by V Guardabasso, C Smith, R Halkes, TV Duong, E Hernandez-Encuentra; comments to author 29.03.22; revised version received 19.04.22; accepted 20.06.22; published 06.07.22.

Please cite as:
Yang K, Hu Y, Qi H
Digital Health Literacy: Bibliometric Analysis
J Med Internet Res 2022;24(7):e35816
URL: https://www.jmir.org/2022/7/e35816
doi:10.2196/35816
PMID:35793141

Original Paper

Use of Multiple Correspondence Analysis and K-means to Explore Associations Between Risk Factors and Likelihood of Colorectal Cancer: Cross-sectional Study

Dídac Florensa[1,2], MPH; Jordi Mateo-Fornés[1], PhD; Francesc Solsona[1], PhD; Teresa Pedrol Aige[3], MPH; Miquel Mesas Julió[2], MPH; Ramon Piñol[4], BSc; Pere Godoy[5,6,7], PhD

[1]Department of Computer Science, University of Lleida, Lleida, Spain
[2]Department of Computer Systems, Santa Maria University Hospital, Lleida, Spain
[3]Hospital-based Cancer Registry, Arnau de Vilanova University Hospital, Lleida, Spain
[4]Catalan Health Service, Department of Health, Lleida, Spain
[5]Biomedical Institute Research of Lleida, Lleida, Spain
[6]Centro de Investigación Biomédica en Red, Madrid, Spain
[7]Santa Maria University Hospital, Population Cancer Registry, Lleida, Spain

Corresponding Author:
Dídac Florensa, MPH
Department of Computer Science
University of Lleida
Jaume II, 69
Lleida, 25001
Spain
Phone: 34 973 70 27 00
Email: didac.florensa@gencat.cat

Abstract

Background: Previous works have shown that risk factors are associated with an increased likelihood of colorectal cancer.

Objective: The purpose of this study was to detect these associations in the region of Lleida (Catalonia) by using multiple correspondence analysis (MCA) and k-means.

Methods: This cross-sectional study was made up of 1083 colorectal cancer episodes between 2012 and 2015, extracted from the population-based cancer registry for the province of Lleida (Spain), the Primary Care Centers database, and the Catalan Health Service Register. The data set included risk factors such as smoking and BMI as well as sociodemographic information and tumor details. The relations between the risk factors and patient characteristics were identified using MCA and k-means.

Results: The combination of these techniques helps to detect clusters of patients with similar risk factors. Risk of death is associated with being elderly and obesity or being overweight. Stage III cancer is associated with people aged ≥65 years and rural/semiurban populations, while younger people were associated with stage 0.

Conclusions: MCA and k-means were significantly useful for detecting associations between risk factors and patient characteristics. These techniques have proven to be effective tools for analyzing the incidence of some factors in colorectal cancer. The outcomes obtained help corroborate suspected trends and stimulate the use of these techniques for finding the association of risk factors with the incidence of other cancers.

(J Med Internet Res 2022;24(7):e29056) doi:10.2196/29056

KEYWORDS

colorectal cancer; cancer registry; multiple correspondence analysis; k-means; risk factors

Introduction

Colorectal cancer is the third most common type of cancer worldwide [1,2]. In Europe, around 250,000 new colorectal cancer cases are diagnosed each year, accounting for around 9% of all malignancies. The rates of this cancer increase with industrialization and urbanization. In general, the evidence shows that the incidence increases in countries where the overall

risk of large bowel cancer is low, while in countries with high incidence, the rate has either stabilized or decreased, particularly among younger age groups [3].

In the province of Lleida (Spain), the population-based cancer registries allow the identification and counting of all incident cases (new cases) diagnosed among the residents of this geographical area [4]. The residents of the Lleida region present lifestyles, risk factors, and work activity, which can be used to determine the specific incidence of certain types of cancer. Nearly half the population of the Lleida province live in rural and semiurban areas. As a consequence, their lifestyle is different from that of the more urban populations in other Catalan provinces [5,6]. Thus, they can present different risk factors and socioeconomic status (SES).

Some studies have demonstrated a higher incidence of colorectal cancer among those with low SES and risk factors such as BMI and smoking. A pooled European cohort study [7] demonstrated that adult weight gain was associated with increased risk of several major cancers. They also concluded that the degree, timing, and duration of being overweight and obesity also seemed to be important. More specifically for colon cancer, Guo et al [8] presented a prospective cohort study in northern China. They concluded that obesity increased the risk of colon cancer in males. Regarding smoking, Mizoue et al [9] presented a report evaluating the association in the Japanese population based on a systematic review of epidemiological evidence. This report concluded that tobacco smoking may increase the risk of colorectal cancer in the Japanese population. However, there is still insufficient epidemiological evidence to demonstrate any clear association with colon cancer. Kim et al [10] studied a possible association between SES and the risk of colorectal cancer in women. Their findings suggested that high SES may protect against colorectal cancer in women. The methodology used in these studies was similar, namely, the multivariate regression analysis.

Recent research has applied the techniques used in this study, but none of these studies were for cancer and risk factors. Ugurlu and Cicek [11] used the multiple correspondence analysis (MCA) method to search for relations in ship collisions [11]. However, the k-means algorithm was more widely used in some cancer aspects. Rustam et al [12] applied this technique to obtain the centroid of each cluster and predict the class of every data point in the validation set. Recently, Ronen et al [13] used k-means as an initial step in a deep learning method to evaluate the colorectal cancer subtypes. K-means allowed the detection of relevant clinical patterns that improved the prediction model. Therefore, the use of MCA and k-means to search for the relationship between risk factors and cancer incidence is a novel method.

Several studies [7-10] have found new associations among risk factors, demographic information, and SES in patients with colorectal cancer. These studies have taken a great effort to analyze and compare risk factors such as obesity, cigarette smoking, and SES in patients with colorectal cancer. They used statistical methods, including Cox regression, Spearman rank correlation coefficient, and multilevel logistic regression to estimate the association between variables. However, none of them used a combination of a statistical method like MCA and an artificial intelligence algorithm such as k-means to search for associations between a group of categorical variables.

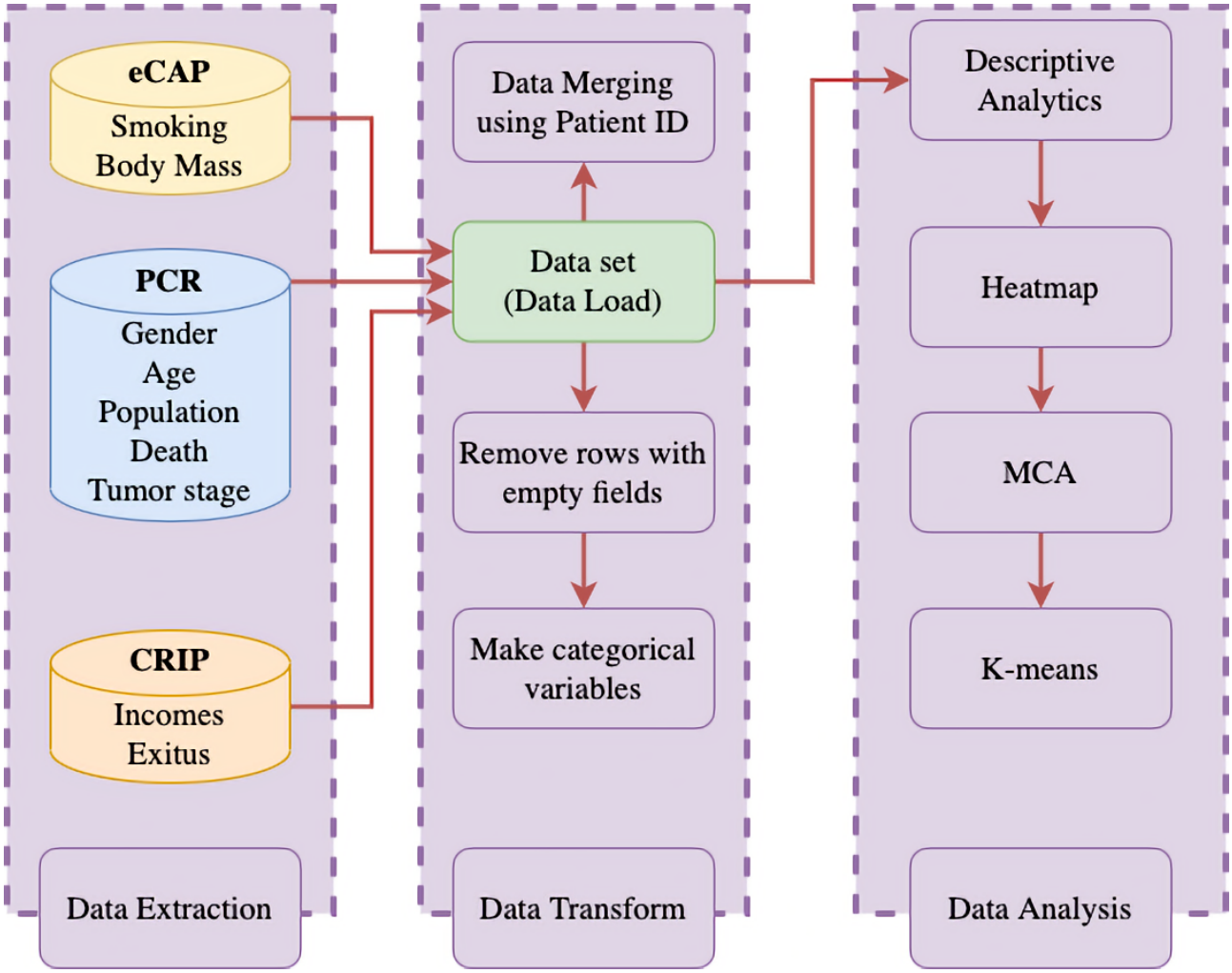
As the main contribution of this study, we propose the use of MCA as a statistical technique to detect relations between risk factors and patients' characteristics and k-means as an unsupervised learning algorithm to search for clusters of patients with similar risk factor profiles for colorectal cancer.

Methods

Preprocessing

The main information sources were the population-based cancer registry of the health region of the province of Lleida, the eCAP (a computerized medical history program used by doctors, pediatricians, and nurses in primary care centers when they see their patients [14]) software, and the Central Register of Insured Persons (a register that allows the unique identification of those covered by the Catalan Health Service through the personal identification code, the management and consultation of their data, and their updates [15]). Before applying the statistical technique, the information was validated by experienced professionals (doctors, nurses, and documentalists) in the Lleida population-based cancer registry who reviewed the clinical history of each patient. After that, the International Agency for Research on Cancer tool was applied to detect unlikely or impossible codes or combinations of codes [16]. Then, an accurate description of the data and basic concepts of the MCA and k-means used in this work are explained in this section. See the system flow chart of the whole process in Figure 1; it shows the different registers used to extract the data, its process and transformation, and its applied analysis. The patients with empty fields were removed.

Figure 1. System flow chart. CRIP: Central Register of Insured Persons; PCR: population-based cancer registry.



Study Population

The colorectal cancer data were extracted from the new cases registered between 2012 and 2015 in the Lleida population-based cancer registry [5,17,18] for patients with cancer in the main hospitals in the health care region of the Lleida province. Specifically, the data set consisted of 1083 new colorectal cancer cases. These hospitals were the Arnau de Vilanova University Hospital and the Santa Maria University Hospital, and the primary information sources were hospital records (International Classification of Diseases, ninth revision codes-140.0 to 208.9) and reports from pathological anatomy. Additionally, these reports confirmed >92% of cases included in the sample. Risk factors such as BMI and smoking were extracted from eCAP software and the SES was extracted from the Central Register of Insured Persons. The study is compliant with the General Data Protection Regulation (European Union), thereby maintaining the anonymity of the patients. Cancer episodes were recorded according to international criteria. In addition, the data analysis (done with R) can be freely downloaded from this GitHub repository [19]. It also included a mock data set randomly generated to test the models. The original data set could not be uploaded due to General Data Protection Regulation, which does not permit sharing patients' information.

The BMI was used to calculate the obesity of each patient by standard weight status categories [20]. We categorized the BMI as the established table: <24.9 as normal weight, 25-29.9 as overweight, and >30 as obesity. Regarding SES, we categorized the variable according to the annual income available from the Central Register of Insured Persons. According to the legislation [21], we created 2 groups: annual income <€18,000 (low income) and >€18,000 (high income) (€1=US $1.04). The population was categorized as rural, semiurban, and urban. In accordance with [22], people living in cities with a population of more than 10,000 were classified as urban, population between 10,000 and 2000 in towns as semiurban, and the rest as rural. The Spanish National Statistics Institute has defined rural areas as those with a population of less than 2000, semiurban areas as those with a population between 2001 and 10,000, and urban areas as those with a population with more than 10,000 people. All the cancer cases that did not conform to one of these fields were discarded automatically. To sum up, each register contains the following fields: age group (50-64 years, 65-74 years, ≥75 years); gender (male, female); population (rural, semiurban, urban); exitus (death, alive); BMI (normal, overweight, obesity); smoking (ex-smoker/smoker, nonsmoker); income (high income, low income); and stage (0, I, II, III, undefined). Table 1 shows the number of cases for each category.

Table 1. Principal comorbidities groups included in this study: patients with colorectal cancer between 2012 and 2015, where all the comorbidities were properly registered (N=1083).

Characteristics	Values, n (%)
Gender	
Male	689 (63.6)
Female	394 (36.4)
Age group (years)	
50-64	319 (29.5)
65-74	328 (30.3)
≥75	436 (40.2)
Exitus	
Death	221 (20.4)
Alive	862 (79.6)
Income[a]	
<€18,000/year	863 (79.7)
>€18,000/year	220 (20.3)
Population	
Rural	228 (21.1)
Semiurban	333 (30.7)
Urban	522 (48.2)
BMI	
Normal	234 (21.6)
Overweight	506 (46.7)
Obesity	343 (31.7)
Smoker	
Smoker/Ex-smoker	232 (21.4)
Nonsmoker	851 (78.6)
Stage	
0	64 (5.9)
I	115 (10.6)
II	168 (15.5)
III	91 (8.4)
Undefined	645 (59.6)

[a]€1=US $1.04.

MCA Algorithm

MCA is an unsupervised learning algorithm for visualizing the patterns in large and multidimensional categorical data [23]. This method can be used to analyze, explore, summarize, and visualize information contained of individuals described by categorical variables [24]. Unlike correspondence analysis (CA), MCA can deal with more than one categorical variable. This is the main advantage of the MCA technique. In our case, MCA was first used to evaluate the relationships between all the features. MCA was then used to evaluate the relationships among population, age, gender, exitus, BMI, smoking, and tumor stage. Associations between features are represented graphically [25]. The graphs aim to visualize the similarities or differences in the profiles simultaneously, identifying those dimensions that contain most of the data variability. Features or their categories close to each other are significantly related statistically.

The factors were interpreted with the help of various statistical coefficients, which complemented each other to provide a better interpretation. The most common and important are inertia, eigenvalue, contribution, and factorial coordinates. Inertia is a measurement of the dispersion of the set of computed distances between points. Analogously, in principal CA, inertia corresponds to the explained variance of dimensions. The

eigenvalue allows the inertia that a specific category produces to be quantified determining a certain percentage relative to the entire set of the active category. The percentage coordinates (x- and y-axis) of the graph enable the category points in a graph to be represented and established. In MCA, the distance between 2 or more categories of different variables can be interpreted in terms of the associations and correlations between these. If 2 categories present high coordinates and are close in space, this means that they tend to be directly associated [26,27]. If 2 categories present high coordinates but are distant from each other (eg, they have opposite signs), this means that they tend to be inversely associated [28,29]. A heatmap was created to help the interpretation of the MCA. This plot used the intensity of the colors to show the level of association between the variables. Our graphs showed the association by the distance between the categories in the MCA plot.

K-means

K-means [30] is a nonsupervised learning algorithm used in data mining and pattern recognition. The algorithm partitions the data set in *k* predefined distinct nonoverlapping subgroups (clusters) where each data point belongs to only one group. It tries to make the intracluster data points as similar as possible while also keeping the clusters as different (far) as possible. It assigns data points to a cluster such that the sum of the squared distance between the data points and the cluster's centroid is at the minimum. The less variation we have within clusters, the more homogeneity (similarity) there is between the data points within the same cluster. The k-means algorithm is composed of the following steps: (1) it places *k* points in the space represented by the patients who are being clustered, (2) it assigns each patient to the group that has the closest centroid, and (3) when all patients have been assigned, it recalculates the positions of the *k* centroids. Steps 2 and 3 are repeated until the centroids no longer move. This produces a separation of the patients into homogenous groups while maximizing heterogeneity across groups. The optimal number of clusters was obtained by the elbow method [31]. This consists of plotting the explained variation as a function of the number of clusters and picking the elbow of the curve as the number of groups to use. To assess internal cluster quality, cluster stability of the optimal solution was computed using Jaccard bootstrap values with 10,000 runs [32].

Statistical Analysis

All the information presented was analyzed using MCA, an extension of CA, and the k-means algorithm. The combination of MCA and k-means benefits the effectiveness of the calculation process and, in consequence, the k-means results. MCA helps to reduce the noise, which allows the k-means algorithm to obtain more accurate distances. The MCA dimension reduction automatically performs data clustering according to the k-means objective function [33]. In addition, the potential confounding factors in this study were assessed by calculating the distances between the variables (inertia) that take into account their relative weight in the database as a whole. However, these variables were related to each other depending on the similarity of each register. Previously, the patients with empty fields were removed.

The MCA method was implemented in scripts performed with R [34], an open-source programming language and environment for statistical computing and graphics. Specifically, the main library used to implement the methods and obtain the results was FactoMineR [35]. K-means was written in Python [36], and the main library used scikit-learn [37]. These methods were launched by their default configuration and using a personal computer.

Results

MCA and K-means Without the Tumor Staging

The analysis of the MCA and k-means without the stage variable included 1083 registers. Figure 2 shows the different categories and their possible associations. The variance for dimension 1 was 15% (eigenvalue 0.21) and that for dimension 2 was 12% (eigenvalue 0.17). Figure 2 also shows the position of each category in the plot and its contribution on the dimensions. Note the contribution of mortality (15% on the negative x-axis and 10.2% on the positive y-axis), the ≥75 years age group (18.8% on the negative x-axis and 4.5% on the positive y-axis), and the ex-smoker/smoker (16.5% and 12.3% on positive x-y axis). Figure 3 shows the relation between the categories. The associations between the points were significant when they were closer and the distance was minimum. For example, females and obesity were represented in the same dimension in the MCA plot. Therefore, the heatmap also demonstrated this association with a distance of 0.4 between the points in the MCA plot.

Figure 2. 2D multiple correspondence analysis plot showing the correlations between the categories and their contributions for all data sets.

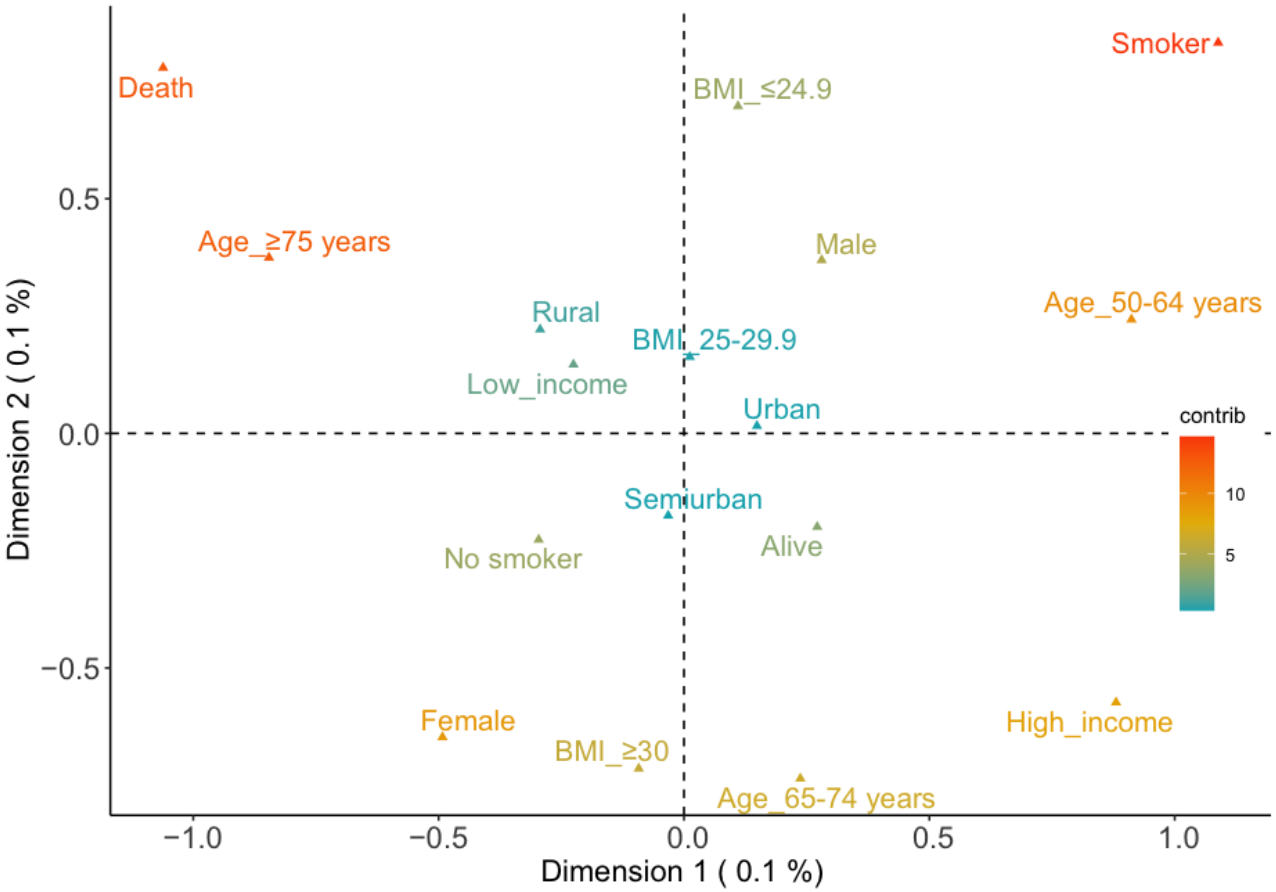
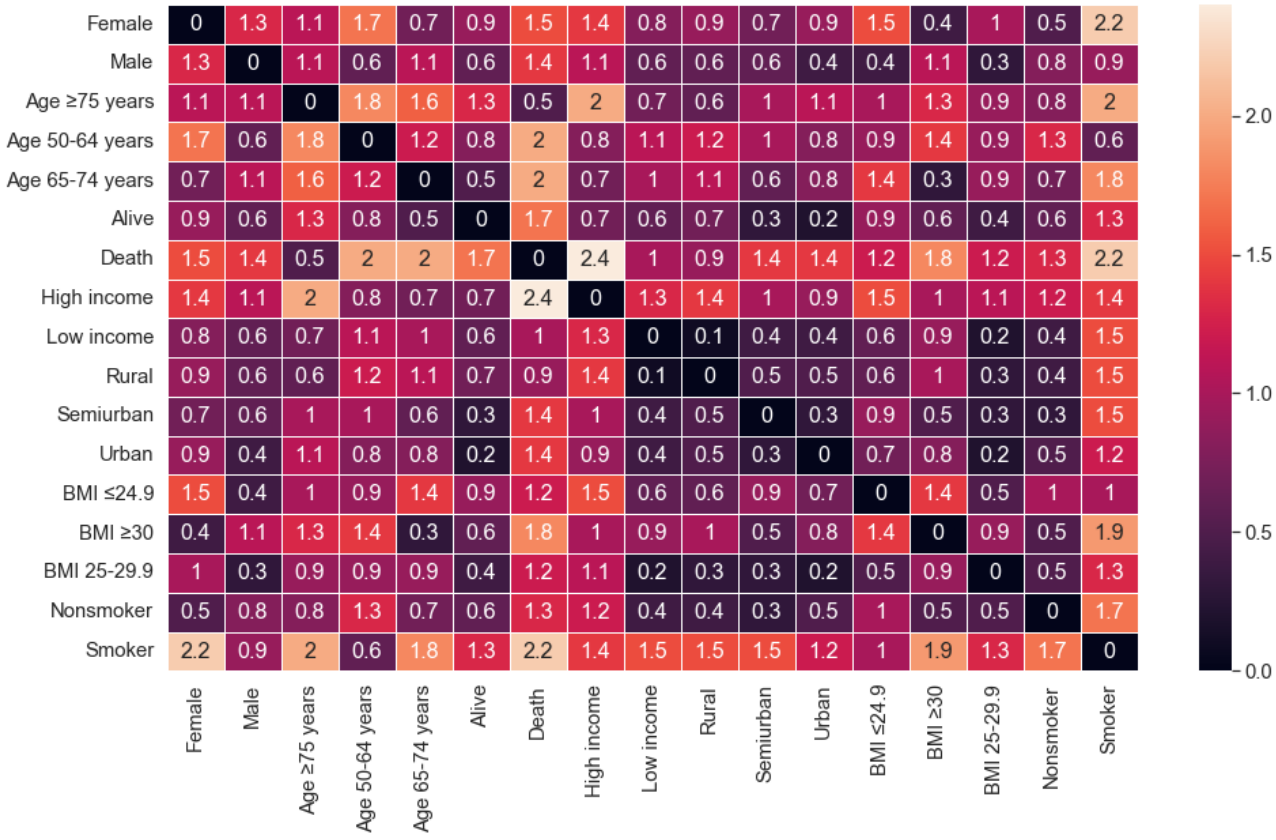


Figure 3. Correlations between the categories by the distance between them.



Graphically, the points closer to each other or the points represented in the same direction of the axis suggest associations. As can be seen, mortality and older age are very close in the plot. This suggests a possible association. Another possible relation observed could be between females and obesity. Then, a cloud on the positive x-axis and the negative y-axis was made up of the 65-74 years age group, high income, and survival. Finally, additional associations could be made up of the 50-64 years age group, males, smokers or ex-smokers, and normal weight.

Table 2 shows the centroids of the main clusters obtained after applying the k-means algorithm. The recommended number of optimal clusters was 5 [31] (see the GitHub [19] repository to evaluate the plot). The first cluster grouped 242 registers among which the main register was males aged ≥75 years from urban populations, with low income, nonsmokers who were overweight, and with a low risk of dying. The next cluster (259 registers) represented females aged between 50 and 64 years with high income. It grouped the cases from rural populations with normal weight and survival. Cluster number 3 was made up of 180 registers. These were mostly males aged ≥75 years with low income and from semiurban populations. They were nonsmokers but were obese and unfortunately included exitus. It was the only cluster that included mortality. The fourth cluster represented urban males aged between 65 and 74 years and with low income. In this case, they were smokers or ex-smokers with normal weight and no mortality. It contained 194 registers. Finally, the last cluster was made up of 208 cases, which included semiurban females aged between 65 and 74 years with low income. They were not smokers but they were overweight. Fortunately, surviving patients predominated in this cluster and the risk of dying was low. See these clusters represented graphically in the GitHub [19].

Table 2. Centroids of the main clusters obtained from the k-means algorithm for all data sets.

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Urban	Rural	Semiurban	Urban	Semiurban
Age ≥75 years	Age 50-64 years	Age ≥75 years	Age 65-74 years	Age 65-74 years
Low income	High income	Low income	Low income	Low income
Male	Female	Male	Male	Female
Nonsmoker	Nonsmoker	Nonsmoker	Smoker/Ex-smoker	Nonsmoker
Overweight	Normal weight	Obesity	Normal weight	Overweight
Alive	Alive	Death	Alive	Alive

MCA and K-means Including the Tumor Staging

This subsection presents the outcomes, including the stage of the tumor. The data set used for this analysis discarded the registers, which did not contain the stage (647 registers). Therefore, the number of cases analyzed was 438 (Table 1). Figure 4 shows the outcomes obtained after applying MCA. The variance of dimension 1 was 11.4% (eigenvalue 0.18) and that of dimension 2 was 10.2% (eigenvalue 0.16). Mortality was also one of those with the highest contribution (26.4% on the positive x-axis and 10.5% on the positive y-axis). Near this was stage III with a high contribution (16.3% on the positive x-axis and 13.7% on the positive y-axis). Ex-smoker/smoker contributed significantly compared with the rest of categories (9.1% on the negative x-axis and 1.3% on the positive y-axis). The relations between these and other categories are shown in Figure 5. See the death and its correlation between stage III. The heatmap differentiated this association clearly, as the MCA plot also showed. The location of the categories in the plot and their contributions suggested possible associations. The main association was between stage III and mortality and with females with stage II, the ≥75 years age group, and nonsmokers. Another relation could be males with high income, aged between 50 and 64 years, stage 0, and ex-smokers or smokers. However, these results could be affected by the decrease in cases.

Figure 4. 2D multiple correspondence analysis plot showing the correlations between the categories and their contributions.

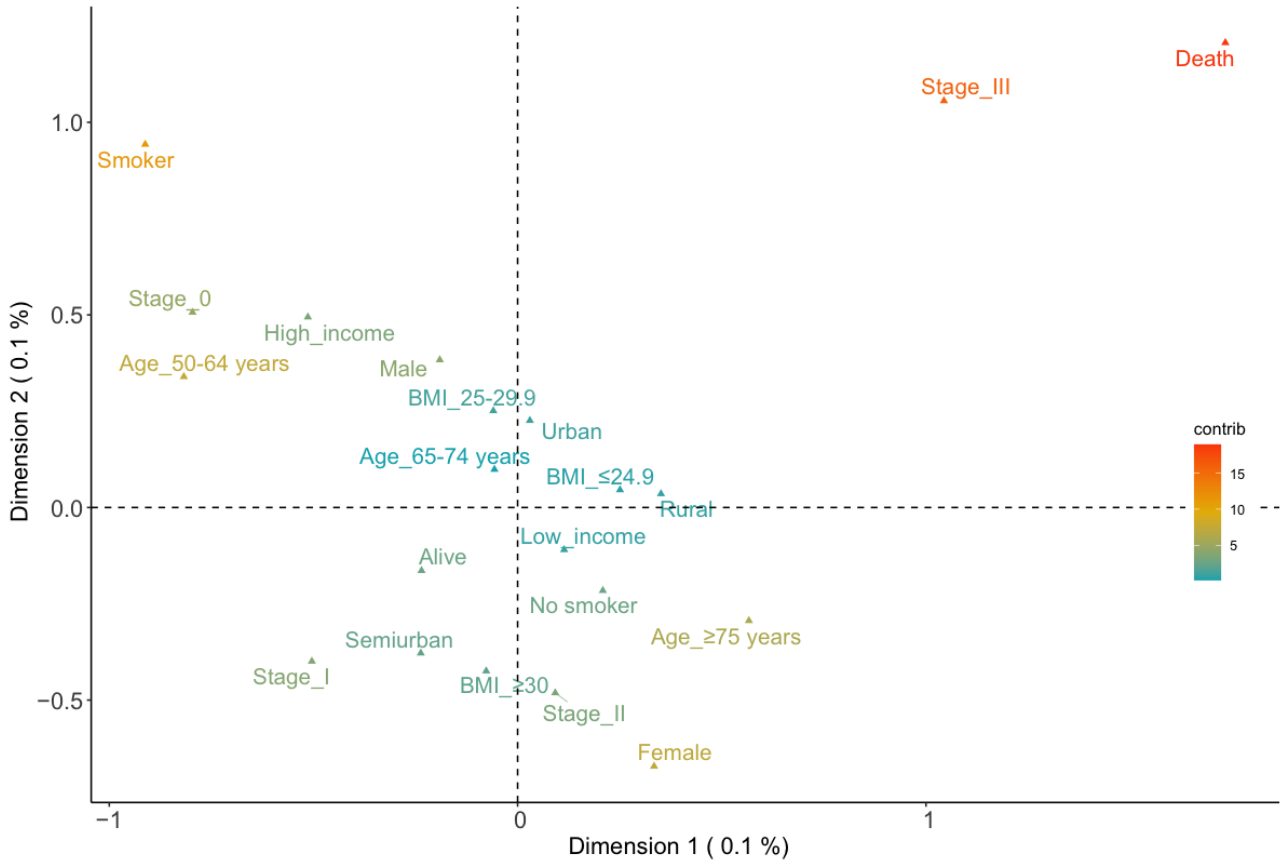


Figure 5. Correlations between the categories by the distance between them including the tumor staging.

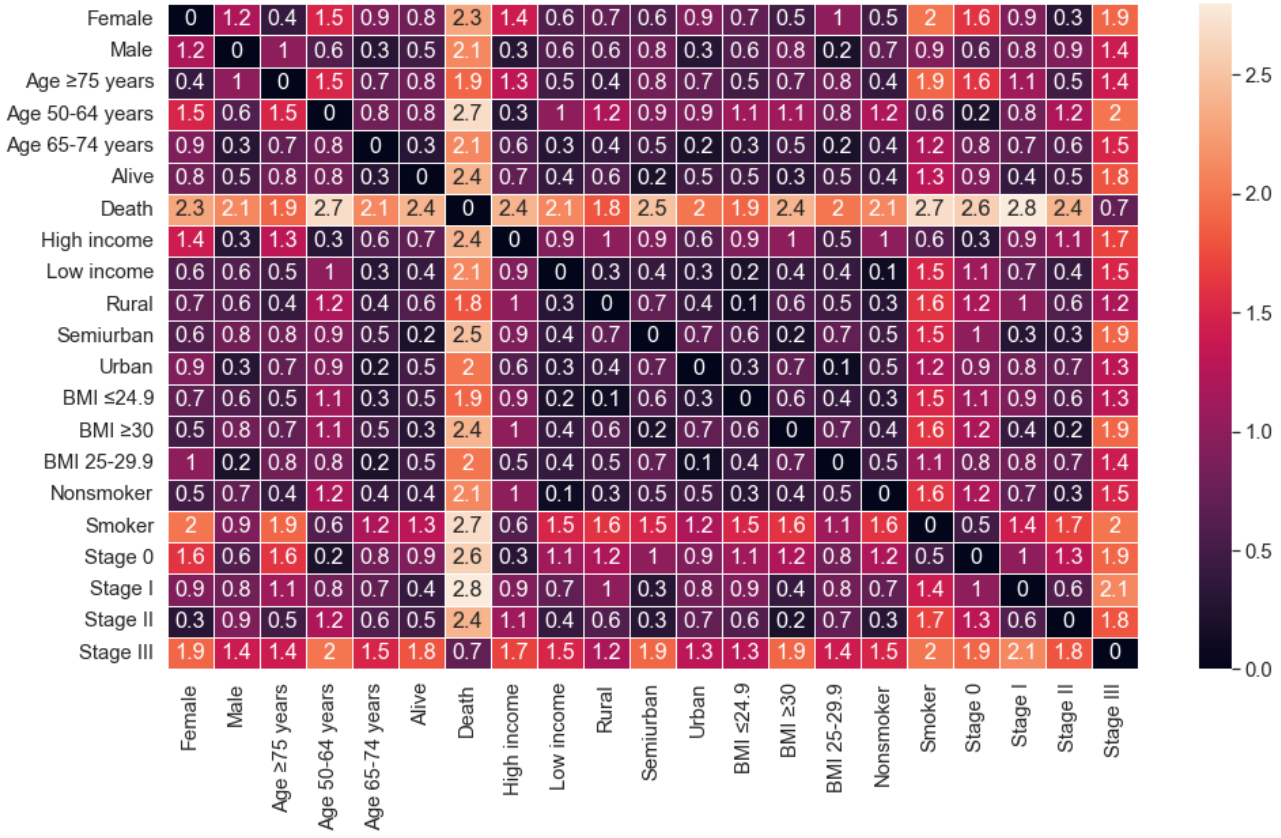


Table 3 shows the clusters obtained from the data set with the tumor stage. All the clusters obtained were male nonsmokers owing to the decrease in the number of registers in the data set. The first cluster with 135 cases represented obese urban patients aged between 65 and 74 years with stage II reached and a low risk of death. The second cluster had 120 registers of patients with stage II and age ≥75 years from semiurban populations. Their risk of death was also low. The next cluster included 76 registers and they were from the urban population but overweight. They included the younger patients (50-64 years age group), with a low risk of death and the lowest stage (stage 0). The fourth cluster (n=72) represented rural inhabitants, aged between 65 and 74 years. They were obese with stage III cancer but low risk of death. However, the fifth cluster was patients from the semiurban population, aged ≥75 years, overweight, in an advanced stage (III), and with a high risk of death. See these clusters represented graphically in the k-means folder of GitHub [19].

Table 3. Centroids of the main clusters obtained from the k-means algorithm: the final data set after including the stage of the tumor.

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Urban	Semiurban	Urban	Rural	Semiurban
Age 65-74 years	Age ≥75 years	Age 50-64 years	Age 65-74 years	Age ≥75 years
High income	Low income	Low income	Low income	Low income
Male	Male	Male	Male	Male
Nonsmoker	Nonsmoker	Nonsmoker	Nonsmoker	Nonsmoker
Obesity	Obesity	Overweight	Obesity	Overweight
Alive	Alive	Alive	Alive	Death
Stage II	Stage II	Stage 0	Stage III	Stage III

Discussion

The MCA technique and the k-means algorithm permit the analysis and detection of clusters of patients with similar risk factors and outcomes not observed in the literature. The population-based cancer registry for the province of Lleida registered 1083 colorectal cancers between 2012 and 2015. This cancer is the most incident in our region [5,17,18] and by applying MCA and k-means, some relationships were found between some aspects that corroborate the usefulness of these techniques. They helped to detect that in colorectal cancer, the age group and BMI risk factors are related. Another important corroboration was the risk of death in older people (≥75 years age group) either obese or overweight and in an advanced stage. Related to this latter factor, the advanced stage was observed in older people with obesity. Stages II and III were 65% (119/181) of the total in the ≥75 years age group.

Previous studies have used clustering techniques to detect associations, but none of them were used for associating patient profiles with risk factors. We based our study on a preliminary paper [38], which evaluated the relationship between air pollution, particulate matter components, and risk of breast cancer in a United States–wide prospective cohort by using a clustering technique. That study concluded that air pollution measures were related to both invasive breast cancer and ductal carcinoma in situ within certain geographic regions. Another starting point was the study presented in [39], which used the combination of MCA and k-means to ascertain multimorbidity patterns. That study concluded that these techniques could help to identify these patterns. Another study our work was based on is the one presented in [40], which studied the trends in the incidence of cancers associated with being overweight and obese. Another study [41] analyzed the possible relation between obesity and colorectal cancer. These papers studied the impact of the risk factors on colorectal cancer but did not use the MCA technique or k-means algorithm to explore associations between these and their impact. In addition, a previous study used MCA to analyze the prognosis in surgery for low rectal cancer [42]. Another study used k-means to search patterns in patients with colorectal cancer, but its main aim was to detect emotion regulation patterns and personal resilience [43]. However, to the best of our knowledge, no prior studies have used MCA or k-means to link types of risk factors, SES, tumor stage, and patients' characteristics in cases of colorectal cancer.

One MCA outcome was the inertia (27%). Further, various variables had high contributions. A strong relation was obtained between older patients (≥75 years age group) and mortality. This may suggest an increase in the risk of mortality for colorectal cancer in older adults, as previous studies showed [44]. On the opposite side of previous associations, it showed another association between survival, high SES, and the 65-75 years age group. Even though the contributions of these are lower than those of mortality and the older population, it is suggested that the risk of death is lower in people with high SES [45] and among younger people. An association was detected between females and obesity although this was not reflected in the k-means. This relation may be because 37% (146/394) of all the women were obese. However, obese men represented 29% (205/689) of the male population, and the percentage of obesity in the data set was 31% (343/1083). This relation suggests that obese women could more likely develop colorectal cancer than men. In general, the probability of colorectal cancer in obese patients can increase by 30%-70% [46]. However, although the contribution is too low to establish a strong relation, the position of males and normal weight in the plot might suggest that there may be some other factors that increase the risk of this cancer and that these techniques highlighted other associations. Some additional patient clinical history would be necessary.

Regarding the k-means analysis, the third cluster confirmed the mortality in the older population with obesity [44]. The first cluster also represented the ≥75 years age group but who were overweight and had no exitus. These differences between clusters suggested that obesity may be a determining factor in older persons that increases the risk of death. In addition, these 2 clusters were males. Similar outcomes were obtained in the fifth cluster when the tumor stage was added. Stage III was directly related with the ≥75 years age group, the semiurban population, and mortality, thereby suggesting that for older persons, being overweight or obese and in an advanced stage could increase the risk of death. The fourth cluster was made up of smokers or ex-smokers. Although tobacco is not usually directly related with colorectal cancer, some studies also support this result [47,48].

The analysis then studied the data set filtered by tumor stage. The final data set was made up of 438 registers. The MCA technique obtained a significant relation between stage III and mortality. However, screening programs and technology decrease this risk, as recent studies concluded [49]. We can also see that stage 0 was related with younger people (50-64 years age group). The k-means results gave similar conclusions as in the MCA. The younger people, stage 0, and survival appeared in the same cluster as demonstrated in the previous k-means analysis with the second cluster. This suggests the importance of screening programs to detect tumors at an early stage [50]. The fourth cluster in the second analysis related rural and stage III. This association may insinuate a possible delay in diagnosis or difficulties in accessing the health care system and mass screening testing in rural areas [51]. Finally, note that all clusters that had stage II or III also included obesity or excess weight. This may suggest that the BMI may be a determinant for having an aggressive colorectal tumor. However, no significant outcomes related to income were obtained, although 80% (863/1083) of the cases were low-income patients. This high percentage of low-income cases could be explained by the fact that the average annual net income per person in Catalonia in 2015 was €12,283 [52].

The strengths of using the MCA and k-means cluster analysis are that the results are less susceptible to outliers in the data, the influence of chosen distance measures, or the inclusion of inappropriate or irrelevant variables [53]. This study had some limitations that should be noted. Regarding the techniques, it tends to take into account the relative weight of each variable concerning the set of study variables and allows control for potential confounding factors such as sex, age, and survival. However, some residual confounding effects cannot be ruled out. Further, these include the low number of cases with tumor stage (438/1083, 40% of total). In consequence, the final data set also made it difficult to analyze the strength of the causal relationship between different prediction parameters and outcomes because it contained few registers. The postal address registered for each case was the patient's home address at the time of cancer diagnosis. However, this address may have changed during the study. Despite this, the number of cases with changed addresses would be very low and this factor is not expected to produce bias in the results. Some lifestyle aspects such as alcohol consumption, diabetes, or profession were not considered. The lack of cause of death is another limitation. The results showed that there is room for other kinds of risk factors. Additional patient clinical history would be required in order to find these. Further, related to the comorbidities, the Charlson index could not be added because approximately only 15% of the sample received it. A future study may be the study of the causality, adding synthetic data to enlarge the data set. Finally, some associations could hide others due to these techniques even though they showed the most significant relationships. In addition, the genetic and hereditary conditions were not considered.

In conclusion, many studies demonstrate that some risk factors such as BMI, tobacco smoking, or SES could influence the incidence of colorectal cancer by using traditional techniques. This study used new techniques such as MCA and k-means to analyze the relationships between colorectal cancer and risk factors. The outcomes obtained demonstrated that the combination of these techniques could help to detect relations between risk factors and patient characteristics. Obesity and being overweight in the older population (≥75 years age group) increases the risk of developing aggressive tumors and death. Stage 0 was related with younger people and survival. This highlights the importance of screening programs for colorectal cancer. The presence of tobacco in a cluster indicated that it must be considered as a risk factor in colorectal cancer. The results of our study help to corroborate suspected trends in several of the relationships detected and confirm the usefulness of these techniques. Further, they encourage applying these methods to other cancers and detecting how the risk factors could be associated. In future work, it is important to delve deeper into the patients' characteristics and risk factors. This means including new variables such as diabetes, alcoholism, or the cause of death. The findings obtained in this study motivate us to search for relations between risk factors in other cancers. Moreover, new techniques and artificial intelligence algorithms can be implemented to explore patterns of pretumor and posttumor detection from the clinical history.

Acknowledgments

This work was supported by contract 2019-DI-43 from the Industrial Doctorate Program of the Government of Catalonia and by the Spanish Ministry of Science and Innovation under contract PID2020-113614RB-C22. Some of the authors are members of the research group 2014-SGR163, funded by the Generalitat de Catalunya.

Conflicts of Interest

None declared.

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 2015 Mar 01;136(5):E359-E386 [FREE Full text] [doi: 10.1002/ijc.29210] [Medline: 25220842]
2. Ferlay J, Colombet M, Soerjomataram I, Dyba T, Randi G, Bettio M, et al. Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. Eur J Cancer 2018 Nov;103:356-387. [doi: 10.1016/j.ejca.2018.07.005] [Medline: 30100160]
3. Labianca R, Beretta GD, Kildani B, Milesi L, Merlin F, Mosconi S, et al. Colon cancer. Crit Rev Oncol Hematol 2010 May;74(2):106-133. [doi: 10.1016/j.critrevonc.2010.01.010] [Medline: 20138539]
4. Parkin DM. The evolution of the population-based cancer registry. Nat Rev Cancer 2006 Aug;6(8):603-612. [doi: 10.1038/nrc1948] [Medline: 16862191]
5. Florensa D, Pedrol T, Mòdol I, Farré X, Salud A, Mateo J, et al. Resultats de l'any 2014. Butlletí Epidemiològic Catalunya 2020 2020;40(12):252-264.
6. Florensa D, Godoy P, Mateo J, Solsona F, Pedrol T, Mesas M, et al. The Use of Multiple Correspondence Analysis to Explore Associations Between Categories of Qualitative Variables and Cancer Incidence. IEEE J Biomed Health Inform 2021 Sep;25(9):3659-3667. [doi: 10.1109/JBHI.2021.3073605] [Medline: 33857006]
7. Bjørge T, Häggström C, Ghaderi S, Nagel G, Manjer J, Tretli S, et al. BMI and weight changes and risk of obesity-related cancers: a pooled European cohort study. Int J Epidemiol 2019 Dec 01;48(6):1872-1885. [doi: 10.1093/ije/dyz188] [Medline: 31566221]
8. Guo L, Li N, Wang G, Su K, Li F, Yang L, et al. [Body mass index and cancer incidence:a prospective cohort study in northern China]. Zhonghua Liu Xing Bing Xue Za Zhi 2014 Mar;35(3):231-236. [Medline: 24831616]
9. Mizoue T, Inoue M, Tanaka K, Tsuji I, Wakai K, Nagata C, Research Group for the Development, Evaluation of Cancer Prevention Strategies in Japan. Tobacco smoking and colorectal cancer risk: an evaluation based on a systematic review of epidemiologic evidence among the Japanese population. Jpn J Clin Oncol 2006 Jan;36(1):25-39. [doi: 10.1093/jjco/hyi207] [Medline: 16423841]
10. Kim D, Masyn KE, Kawachi I, Laden F, Colditz GA. Neighborhood socioeconomic status and behavioral pathways to risks of colon and rectal cancer in women. Cancer 2010 Sep 01;116(17):4187-4196 [FREE Full text] [doi: 10.1002/cncr.25195] [Medline: 20544839]
11. Ugurlu H, Cicek I. Analysis and assessment of ship collision accidents using Fault Tree and Multiple Correspondence Analysis. Ocean Engineering 2022 Feb;245:110514. [doi: 10.1016/j.oceaneng.2021.110514]
12. Rustam Z, Hartini S, Yunus R, Pratama R, Yunus R, Hidayat R. Analysis of Architecture Combining Convolutional Neural Network (CNN) and Kernel K-Means Clustering for Lung Cancer Diagnosis. Artic Int J Adv Sci Eng Inf Technol Internet 2020 2020;10(3):1200-1206. [doi: 10.18517/ijaseit.10.3.12113]
13. Ronen J, Hayat S, Akalin A. Evaluation of colorectal cancer subtypes and cell lines using deep learning. Life Sci Alliance 2019 Dec;2(6):e201900517 [FREE Full text] [doi: 10.26508/lsa.201900517] [Medline: 31792061]
14. eCAP. Departament de Salut. URL: https://salutweb.gencat.cat/ca/ambits_actuacio/linies_dactuacio/tic/sistemes-informacio/gestio-assistencial/ecap/ [accessed 2020-12-16]
15. Registre central de població del CatSalut. Català de la Salut. URL: https://catsalut.gencat.cat/ca/proveidors-professionals/registres-catalegs/registres/central-poblacio/index.html [accessed 2020-12-16]
16. International Agency for Research on Cancer. URL: http://www.iacr.com.fr/index.php?option=com_content&view=article&id=72:iarccrgtools&catid=68&Itemid=445 [accessed 2020-12-14]
17. Godoy P, Pedrol T, Mòdol I, Salud A. El Registre poblacional de càncer a Lleida: resultats i perspectives. Butlletí Epidemiològic Catalunya. 2016. URL: https://scientiasalut.gencat.cat/handle/11351/3052?show=full&locale-attribute=en [accessed 2022-07-06]
18. Godoy-Garcia P, Pedrol T, Mòdol-Pena I, Salud A. El registre poblacional de càncer a Lleida: resultats de l'any 2013. Butlletí Epidemiològic Catalunya. 2018. URL: https://scientiasalut.gencat.cat/handle/11351/3665?show=full [accessed 2022-07-06]
19. Florensa D, Godoy P, Solsona P, et al. The use of multiple correspondence analysis to explore associations between categories of qualitative variables and cancer incidence. Github Repository. URL: https://github.com/didacflorensa/MCA-Cancer [accessed 2021-01-01]
20. Weisell R. Body mass index as an indicator of obesity. Asia Pac J Clin Nutr 2002;11(8):681-684. [doi: 10.1046/j.1440-6047.11.s8.5.x]
21. España, Real Decreto-ley 16/2012, de 20 de abril, de medidas urgentes para garantizar la sostenibilidad del Sistema Nacional de Salud y mejorar la calidad y seguridad de sus prestaciones. Boletín Oficial del Estado. URL: https://www.boe.es/diario_boe/txt.php?id=BOE-A-2012-5403 [accessed 2022-07-06]
22. García GJ. La Población Rural de España: De los Desequilibrios a la Sostenibilidad Social. Spain: Barcelona: Fundación La Caixa; 2013:146-149.
23. Murtagh F. Multiple correspondence analysis and related methods. Psychometrika 2007 Mar 24;72(2):275-277. [doi: 10.1007/s11336-006-1579-x]

24. Husson F, Josse J. Multiple correspondence analysis. In: Visualization and Verbalization of Data. Boca Raton, Florida: CRC/PRESS; 2014.
25. Sourial N, Wolfson C, Zhu B, Quail J, Fletcher J, Karunananthan S, et al. Correspondence analysis is a useful tool to uncover the relationships among categorical variables. J Clin Epidemiol 2010 Jun;63(6):638-646 [FREE Full text] [doi: 10.1016/j.jclinepi.2009.08.008] [Medline: 19896800]
26. Greenacre M. Correspondence Analysis in Practice, Third Edition. Boca Raton, Florida: Chapman and Hall/CRC; 2017.
27. Roux BL, Rouanet H. Geometric Data Analysis: From Correspondence Analysis to Structured Data Analysis. New York: Springer; 2005.
28. Di Franco G. Multiple correspondence analysis: one only or several techniques? Qual Quant 2015 Apr 21;50(3):1299-1315. [doi: 10.1007/s11135-015-0206-0]
29. Heckler CE. Applied Multivariate Statistical Analysis. Technometrics 2005 Nov;47(4):517-517. [doi: 10.1198/tech.2005.s319]
30. Likas A, Vlassis N, J. Verbeek J. The global k-means clustering algorithm. Pattern Recognition 2003 Feb;36(2):451-461. [doi: 10.1016/s0031-3203(02)00060-2]
31. Bholowalia P, Kumar A. EBK-Means: A Clustering Technique based on Elbow MethodK-Means in WSN. Int J Comput Appl 2014;105:17-24. [doi: 10.5120/18405-9674]
32. Hennig C. Cluster-wise assessment of cluster stability. Computational Statistics & Data Analysis 2007 Sep;52(1):258-271. [doi: 10.1016/j.csda.2006.11.025]
33. Ding C, He X. K-means clustering via principal component analysis. 2004 Presented at: Proceedings of the 21st Int Conf Mach Learn (ICML); 2004; Banff, Canada. [doi: 10.1145/1015330.1015408]
34. The R Project for Statistical Computing. 2019. URL: https://www.r-project.org/ [accessed 2020-12-16]
35. Lê S, Josse J, Husson F. FactoMineR: An R Package for Multivariate Analysis. J Stat Soft 2008;25(1):1-18. [doi: 10.18637/jss.v025.i01]
36. Welcome to Python.org. URL: https://www.python.org/ [accessed 2020-12-16]
37. scikit-learn: MLIP. URL: https://scikit-learn.org/stable/ [accessed 2020-12-16]
38. White A, Keller J, Zhao S, Carroll R, Kaufman J, Sandler D. Air Pollution, Clustering of Particulate Matter Components, and Breast Cancer in the Sister Study: A U.S.-Wide Cohort. Environ Health Perspective 2019;127(10):107002. [doi: 10.1289/ehp5131]
39. Violán C, Roso-Llorach A, Foguet-Boreu Q, Guisado-Clavero M, Pons-Vigués M, Pujol-Ribera E, et al. Multimorbidity patterns with K-means nonhierarchical cluster analysis. BMC Fam Pract 2018 Jul 03;19(1):108 [FREE Full text] [doi: 10.1186/s12875-018-0790-x] [Medline: 29969997]
40. Steele CB, Thomas CC, Henley SJ, Massetti GM, Galuska DA, Agurs-Collins T, et al. Vital Signs: Trends in Incidence of Cancers Associated with Overweight and Obesity - United States, 2005-2014. MMWR Morb Mortal Wkly Rep 2017 Oct 03;66(39):1052-1058 [FREE Full text] [doi: 10.15585/mmwr.mm6639e1] [Medline: 28981482]
41. Liu P, Wu K, Ng K, Zauber AG, Nguyen LH, Song M, et al. Association of Obesity With Risk of Early-Onset Colorectal Cancer Among Women. JAMA Oncol 2019 Jan 01;5(1):37-44 [FREE Full text] [doi: 10.1001/jamaoncol.2018.4280] [Medline: 30326010]
42. Mancini R, Pattaro G, Diodoro MG, Sperduti I, Garufi C, Stigliano V, et al. Tumor Regression Grade After Neoadjuvant Chemoradiation and Surgery for Low Rectal Cancer Evaluated by Multiple Correspondence Analysis: Ten Years as Minimum Follow-up. Clin Colorectal Cancer 2018 Mar;17(1):e13-e19. [doi: 10.1016/j.clcc.2017.06.004] [Medline: 28865674]
43. Baziliansky S, Cohen M. Emotion Regulation Patterns among Colorectal Cancer Survivors: Clustering and Associations with Personal Coping Resources. Behav Med 2021;47(3):214-224. [doi: 10.1080/08964289.2020.1731674] [Medline: 32275195]
44. Liu X, Bi Y, Wang H, Meng R, Zhou W, Zhang G, et al. Different trends in colorectal cancer mortality between age groups in China: an age-period-cohort and joinpoint analysis. Public Health 2019 Jan;166:45-52. [doi: 10.1016/j.puhe.2018.08.007] [Medline: 30447645]
45. Hastert TA, Beresford SAA, Sheppard L, White E. Disparities in cancer incidence and mortality by area-level socioeconomic status: a multilevel analysis. J Epidemiol Community Health 2015 Mar;69(2):168-176. [doi: 10.1136/jech-2014-204417] [Medline: 25288143]
46. Bardou M, Barkun AN, Martel M. Obesity and colorectal cancer. Gut 2013 Jun;62(6):933-947. [doi: 10.1136/gutjnl-2013-304701] [Medline: 23481261]
47. Rawla P, Sunkara T, Barsouk A. Incidence, mortality, survival, and risk factors. Gastroenterology Review 2019;14(2):89-103. [doi: 10.5114/pg.2018.81072]
48. Botteri E, Iodice S, Bagnardi V, Raimondi S, Lowenfels AB, Maisonneuve P. Smoking and colorectal cancer: a meta-analysis. JAMA 2008 Dec 17;300(23):2765-2778. [doi: 10.1001/jama.2008.839] [Medline: 19088354]
49. Brouwer NPM, Bos ACRK, Lemmens VEPP, Tanis PJ, Hugen N, Nagtegaal ID, et al. An overview of 25 years of incidence, treatment and outcome of colorectal cancer patients. Int J Cancer 2018 Dec 01;143(11):2758-2766 [FREE Full text] [doi: 10.1002/ijc.31785] [Medline: 30095162]

50. Brenner H, Chang-Claude J, Jansen L, Knebel P, Stock C, Hoffmeister M. Reduced risk of colorectal cancer up to 10 years after screening, surveillance, or diagnostic colonoscopy. Gastroenterology 2014 Mar;146(3):709-717. [doi: 10.1053/j.gastro.2013.09.001] [Medline: 24012982]
51. Henley SJ, Anderson RN, Thomas CC, Massetti GM, Peaker B, Richardson LC. Invasive Cancer Incidence, 2004-2013, and Deaths, 2006-2015, in Nonmetropolitan and Metropolitan Counties - United States. MMWR Surveill Summ 2017 Jul 07;66(14):1-13 [FREE Full text] [doi: 10.15585/mmwr.ss6614a1] [Medline: 28683054]
52. Average annual net income of households. Statistical Institute of Catalonia. 2015. URL: https://www.idescat.cat/pub/?id=aec&n=414&t=2015&lang=en [accessed 2020-12-16]
53. Liao M, Li Y, Kianifard F, Obi E, Arcona S. Cluster analysis and its application to healthcare claims data: a study of end-stage renal disease patients who initiated hemodialysis. BMC Nephrol 2016 Mar 02;17:25 [FREE Full text] [doi: 10.1186/s12882-016-0238-2] [Medline: 26936756]

Abbreviations

CA: correspondence analysis
MCA: multiple correspondence analysis
SES: socioeconomic status

Edited by A Mavragani; submitted 25.03.21; peer-reviewed by CM Moore, E Mohammadi, Y Chu, L Espinosa-Leal; comments to author 28.12.21; revised version received 22.02.22; accepted 23.05.22; published 19.07.22.

Please cite as:
Florensa D, Mateo-Fornés J, Solsona F, Pedrol Aige T, Mesas Julió M, Piñol R, Godoy P
Use of Multiple Correspondence Analysis and K-means to Explore Associations Between Risk Factors and Likelihood of Colorectal Cancer: Cross-sectional Study
J Med Internet Res 2022;24(7):e29056
URL: https://www.jmir.org/2022/7/e29056
doi:10.2196/29056
PMID:35852835

Original Paper

A Technology-Assisted Telephone Intervention for Work-Related Stress Management: Pilot Randomized Controlled Trial

Salla Tuulikki Muuraiskangas[1], MHSc; Anita Marianne Honka[2], MSc; Ulla-Maija Junno[3], MHSc; Hannu Olavi Nieminen[4,5], MSc, DTech; Jouni Kalevi Kaartinen[1], MSc

[1]Smart Health, VTT Technical Research Centre of Finland Ltd, Oulu, Finland

[2]Smart Health, VTT Technical Research Centre of Finland Ltd, Tampere, Finland

[3]Luona Hoiva Ltd, Helsinki, Finland

[4]Movendos Ltd, Tampere, Finland

[5]Faculty of Medicine and Health Technology, Tampere university, Tampere, Finland

Corresponding Author:
Salla Tuulikki Muuraiskangas, MHSc
Smart Health
VTT Technical Research Centre of Finland Ltd
Kaitoväylä 1
Oulu, 90590
Finland
Phone: 358 408350169
Email: salla.muuraiskangas@gmail.com

Abstract

Background: Stress management interventions combining technology with human involvement have the potential to improve the cost-effectiveness of solely human-delivered interventions, but few randomized controlled trials exist for assessing the cost-effectiveness of technology-assisted human interventions.

Objective: The aim of this study was to investigate whether a technology-assisted telephone intervention for stress management is feasible for increasing mental well-being or decreasing the time use of coaches (as an approximation of intervention cost) while maintaining participants' adherence and satisfaction compared with traditional telephone coaching.

Methods: A 2-arm, pilot randomized controlled trial of 9 months for stress management (4-month intensive and 5-month maintenance phases) was conducted. Participants were recruited on the web through a regional occupational health care provider and randomized equally to a research (technology-assisted telephone intervention) and a control (traditional telephone intervention) group. The coaching methodology was based on habit formation, motivational interviewing, and the transtheoretical model. For the research group, technology supported both coaches and participants in identifying behavior change targets, setting the initial coaching plan, monitoring progress, and communication. The pilot outcome was intervention feasibility, measured primarily by self-assessed mental well-being (WorkOptimum index) and self-reported time use of coaches and secondarily by participants' adherence and satisfaction.

Results: A total of 49 eligible participants were randomized to the research (n=24) and control (n=25) groups. Most participants were middle-aged (mean 46.26, SD 9.74 years) and female (47/49, 96%). Mental well-being improved significantly in both groups (WorkOptimum from "at risk" to "good" Â>0.85; *P*<.001), and no between-group differences were observed in the end (Â=0.56, 95% CI 0.37-0.74; *P*=.56). The total time use of coaches did not differ significantly between the groups (366.0 vs 343.0 minutes, Â=0.60, 95% CI 0.33-0.85; *P*=.48). Regarding adherence, the dropout rate was 13% (3/24) and 24% (6/25), and the mean adherence rate to coaching calls was 92% and 86% for the research and control groups, respectively; the frequency of performing coaching tasks was similar for both groups after both phases; and the diligence in performing the tasks during the intensive phase was better for the research group (5.0 vs 4.0, Â=0.58, 95% CI 0.51-0.65; *P*=.03), but no difference was observed during the maintenance phase. Satisfaction was higher in the research group during the intensive phase (5.0 vs 4.0, Â=0.66, 95% CI 0.58-0.73; *P*<.001) but not during the maintenance phase.

Conclusions: The technology-assisted telephone intervention is feasible with some modifications, as it had similar preliminary effectiveness as the traditional telephone intervention, and the participants had better satisfaction with and similar or better

adherence to the intervention, but it did not reduce the time use of coaches. The technology should be improved to provide more digested information for action planning and templates for messaging.

Trial Registration: ClinicalTrials.gov NCT02445950; https://www.clinicaltrials.gov/ct2/show/NCT02445950

(J Med Internet Res 2022;24(7):e26569) doi:10.2196/26569

KEYWORDS

health behavior change intervention; telephone coaching; technology-assisted coaching; remote coaching; occupational health; mental well-being; stress management; feasibility; randomized controlled trial

Introduction

Background

Work-related stress and its indirect consequences for physical, mental, and social well-being are serious threats to public health. Long-term stress increases the risk of sleep problems [1,2], coronary heart disease [3,4], and various mental health problems [5], such as burnout [6,7], anxiety [8], and depression [9,10]. Stress is also linked to an increased number of sick days [11-15], lost productivity, inability to work, and presenteeism at work [16-18]. The estimated stress-related health, economic, and social costs are considerable [13,14,19]. Hence, reducing work-related stress effectively can have a major positive impact on individuals, companies, and society as a whole.

Various interventions have been developed for stress management and mental well-being. Traditionally, interventions have been delivered by human coaches or therapists through face-to-face or telephone sessions, but such interventions are not easily scalable. Nowadays, interventions are often supported by technology, or they can even be delivered fully and automatically by technology without human involvement. The use of technology can decrease human involvement and thus costs [20], which enables scaling up the intervention for a larger population, and therefore, it holds a promise for a healthier population.

Stress management interventions that blend technology and human effort have been found in several studies to be effective at reducing stress compared with no treatment [21-23]. In addition, there is evidence from the mental health perspective that blended interventions are more effective at reducing symptoms of anxiety, distress, and negative thoughts than, for example, fully human interventions [24,25]. Fully automatized interventions without human involvement have also been found to be effective in reducing stress [23,26-30]. However, there are studies indicating that blended stress management interventions can be more effective at reducing stress or stress-related mental health problems than fully technological interventions [31,32].

The level of human and technology involvement and the terminology used to describe the blend varies in health intervention studies [20,33]. In this study, we focus on *technology-assisted human interventions*, where the intervention is produced by a human (therapist, coach, etc) with the assistance of technology. Technology assists the human by facilitating the coaching process or by providing additional insights regarding the participant via automatic analyses of electronic questionnaire responses or passively collected data (eg, via wearables). An example of this kind of intervention is telephone coaching, where technology provides additional insights (eg, participants' self-monitoring data or intervention component suggestions) to support the decision-making of coaches [23,34].

There are only a few randomized controlled trials (RCTs) studying the cost-effectiveness of blended stress management interventions. These studies have shown blended stress management interventions to have an acceptable likelihood for cost-effectiveness compared with waiting list control [35-37], fully human interventions with no technology involved, [38] and fully technological interventions [39]. In addition, blended interventions for mental health are cost-effective compared with fully human interventions, at least from the care provider perspective if not from the societal perspective [40]. In these studies, cost-effectiveness was calculated by comparing medical and societal costs to the measured health improvements [36-38,40] or to the number of patients with a symptom-free status [35]. For instance, cost-effectiveness was evaluated based on the total therapist time spent compared with the measured health improvements in the study by Kaldo et al [38].

Participants' intervention adherence is an important determinant of intervention effectiveness [41] and feasibility, as it defines participants' exposure to the intervention. It is necessary to distinguish between at least two types of adherence, the adherence to the research study and its questionnaires (conversely called dropout attrition) and adherence to the intervention itself, that is, to what extent the components of the intervention are being used (conversely called nonuse attrition) [42]. Previous studies on mental health suggest that intervention adherence (both regarding the study and the use) is higher when human guidance is involved in the intervention [20,41,43-45].

Participants' satisfaction with the intervention is important for intervention feasibility, and it is associated with sustained adherence [46]. Therefore, studying satisfaction sheds light on understanding intervention adherence and effectiveness. The scientific publications studying participant satisfaction in blended stress management interventions are scarce. Earlier studies on mental health [47] have shown that in interventions in which the technology is used more and human assistance is less, the participants are less satisfied than in fully human interventions [47], but there is also evidence of equal satisfaction among the participants in blended interventions and group-based, fully human interventions [48].

In summary, previous research suggests that blended stress management interventions have the potential to be effective, but cost-effectiveness studies are lacking. Furthermore,

adherence and satisfaction are important for evaluating the feasibility of interventions in more detail and for helping to refine their implementation for future large-scale RCTs.

Objectives

The primary objective of this study was to investigate whether a technology-assisted telephone intervention for stress management is feasible for increasing participants' well-being or decreasing the time use of coaches while maintaining participants' adherence and satisfaction compared with a traditional telephone intervention (without technology assistance) in an occupational health care setting. The primary trial outcomes were mental well-being and time use of coaches as an approximation of the intervention cost (ClinicalTrials.gov NCT02445950). As primary analyses, we assessed whether the participants in the research group (technology-assisted telephone intervention) reported a greater improvement in well-being, measured by the WorkOptimum index [49,50], and whether the time use of coaches was lower than that of the control group (telephone intervention). As secondary outcomes, we assessed participants' adherence and satisfaction in both groups. The secondary outcomes mentioned in the trial registration related to the evaluation of the used technology (usefulness, ease of use, and accuracy) will be reported in another paper (Honka, MSc, unpublished data, December 2021).

Methods

Trial Design

A nonblinded, parallel-group, 2-arm pilot RCT was conducted for 9 months in Oulu, Finland, to explore whether a technology-assisted telephone intervention for stress management is feasible for increasing mental well-being or decreasing the time use of coaches while maintaining adherence and satisfaction. The trial registration opened in November 2014, and the trial started in February 2015 and ended in October 2015.

Participants

Participants were recruited from among the employees of the City of Oulu, Finland, via the channels of the regional occupational health care provider. The recruitment announcement was published on the intranet pages of the City of Oulu and the occupational health care provider, and in the magazines of the City of Oulu (to staff) and the occupational health care provider (to customers). The staff of the occupational health care provider also recruited participants personally and via email. The registration of the study was conducted on the web via a link in the announcement. Registered employees received informed consent through regular mail, where information regarding the 2 study groups was provided: intervention, data collection, data processing, data privacy, research partners, and contact details. Signed consent was collected by a research partner who provided the coaching service for the intervention. An electronic eligibility survey was sent to the employees who returned signed consent forms, after which they were informed whether they were accepted to the study or not. The eligibility criteria are presented in Textbox 1. As the quality of romantic relationships substantially influences mental well-being [51], this was selected as one of the intervention areas; therefore, being in a relationship was part of the inclusion criteria.

Textbox 1. Eligibility criteria for the study.

Inclusion criteria

- Own assessment of decreased psychophysical state (based on a subset of items of the WorkOptimum questionnaire)
- Customers of the occupational health care provider who work full-time for the City of Oulu (in the area of information technology, education, culture, social, health, and customer service)
- Age >18 years
- In a relationship, motivated to enhance own well-being by making lifestyle changes or performing exercises related to mental well-being or relationships (based on 1 question in the eligibility survey)

Exclusion criteria

- Night shifts included in the work schedule
- Acute health condition or a serious disease
- Chronic pain affecting physical function
- Long period of absence (eg, long vacation, alternation leave, parental leave, or pension) from work during the intervention period
- Participation in other studies

Randomization

The eligible participants were randomly allocated to either a research (technology-assisted telephone intervention) or control (traditional telephone intervention) group in a 1:1 ratio through stratified block randomization. Group allocation was stratified based on socioeconomic status and having minors as family members, since these factors were anticipated to influence the mental well-being and adherence outcomes of the study owing to challenges in meeting the demands of work and family responsibilities [52-54]. Socioeconomic status was categorized as "lower-level employees" (eg, secretaries, nurses, childminders, and customer servants) or "upper-level employees" (eg, managers, doctors, psychologists, teachers, and information technology professionals) based on job descriptions and the required education level [55]. Having

minors in the family was described using 3 categories: “no children,” “at least one child below school age (<7 years),” and “only school-aged children.”

The participants were randomized simultaneously into the 2 groups via Microsoft Excel (version 2010) using its random number generator. The randomization was conducted by a researcher who was not involved in the study as an investigator. The study investigators were aware of the group to which each participant belonged.

Sample Size

Available coaching resources defined the number of participants that could be enrolled in the study. At the time of the study, the participating coaching service provider used 3 coaches who could use an average of 20% of their time for the study participants. Therefore, the objective was to have 40 participants in the study. As a dropout rate of 20% is common in telephone interventions [34,56], the aim was to recruit 50 participants.

Interventions

Common Intervention Components

There were 2 interventions: a technology-assisted telephone intervention and a traditional telephone intervention. The interventions lasted for 9 months, and they were divided into a 4-month intensive phase and a 5-month maintenance phase (Figure 1).

Coaching was performed by 3 coaches recruited by the research partner, Mawell Care Limited, so that each of them had an equal number of clients from both the control and research groups. The participants were allocated to the coaches based on mutual availability for the first telephone call. Coaching is based on the habit formation theory, according to which small, regularly repeating behavioral actions or tasks support long-term behavior change [57]. The tasks were related to different topics relevant to well-being, namely sleep, physical activity, eating, alcohol consumption, smoking, recovery from stress, anxiety, personal values, workload management, quality of relationship, self-esteem, and weight management. The tasks could be, for example, “take at least 7000 steps a day,” “practice relaxation with the help of audio exercises,” and “keep a diary of personal eating habits for some days.” Some of these tasks included links to external web resources, such as “read information about the consequences of sleep deprivation from [webpage],” and “practice mindful eating with Oiva at a weekly basis.” Oiva is a web portal that includes exercises for mental well-being, which are based on Acceptance and Commitment Therapy [58]. Oiva contains short few-minute exercises under the themes of well-being of mind and body, values, and everyday choices. The number of study participants per study group, working on different task areas, is presented in Table 1.

Behavioral strategies were based on motivational interviewing and the transtheoretical model of Prochaska et al [59,60]. The interventions included several behavior change techniques, as categorized in the study by Michie et al [61]. The applied behavior change techniques are presented in Table 2.

The main difference between the interventions was the number of telephone calls and the use of technology in coaching. The research group had 5 coaching calls during the intensive phase and 1 at the end. The control group had 5 coaching calls in the intensive phase and 3 in the maintenance phase.

Figure 1. Intervention timeline with intervention components. HRS: health recommender system.

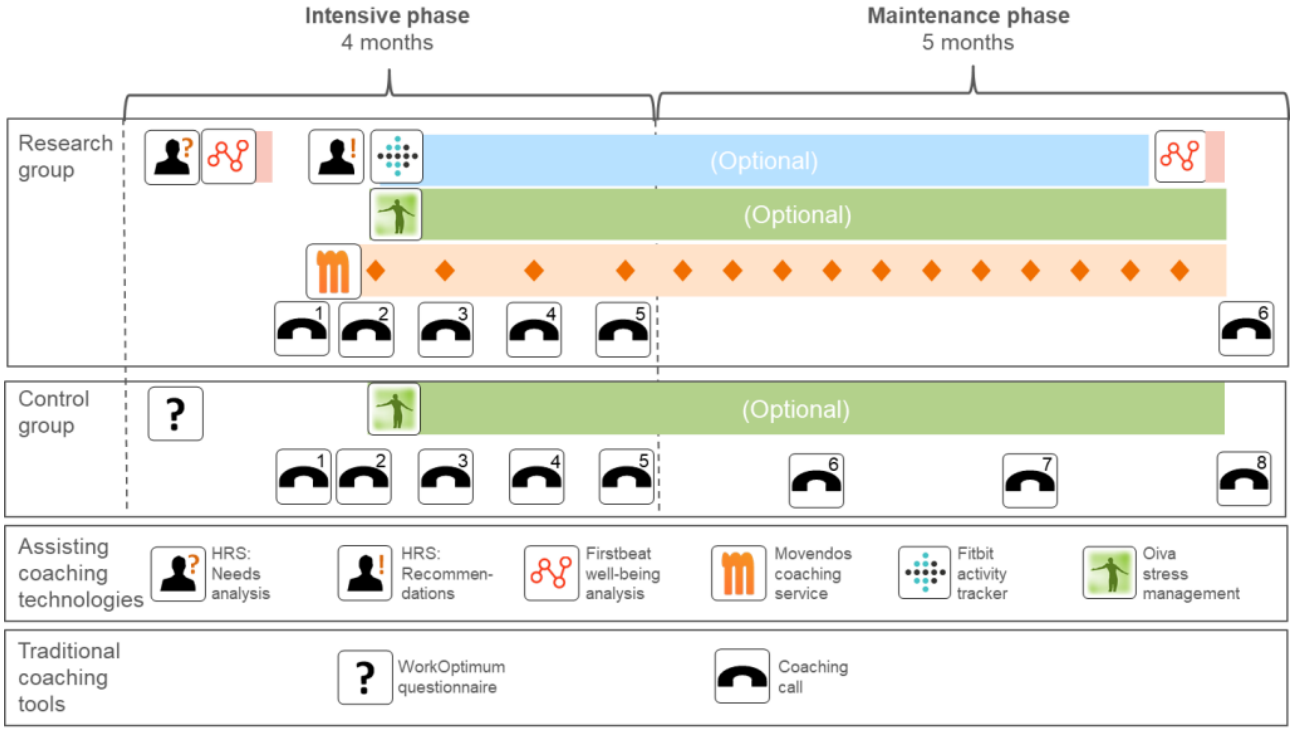


Table 1. Task areas and selection frequency for both groups.

Task areas	Group, n[a] (%)		Total, n (%)
	Research	Control	
Sleep	7 (58)	5 (42)	12 (100)
Physical activity	23 (55)	19 (45)	42 (100)
Eating	9 (47)	10 (53)	19 (100)
Alcohol consumption	1 (50)	1 (50)	2 (100)
Smoking	1 (50)	1 (50)	2 (100)
Recovery from stress, anxiety, or personal values	20 (63)	12 (37)	32 (100)
Workload management	9 (53)	8 (47)	17 (100)
Quality of relationship	3 (100)	0 (0)	3 (100)
Self-esteem	3 (100)	0 (0)	3 (100)
Weight management	5 (45)	6 (55)	11 (100)

[a]Number of participants in a group who selected a task from a specific area.

Table 2. Behavior change techniques used in both interventions.

Phase of the intervention	Behavior change technique
Beginning	• Goal setting of behavior • Goal setting of outcome • Problem solving • Action planning • Behavioral contact • Information about health consequences • Pros and cons • Comparative imagining of future outcomes
Coaching during intensive and maintenance phases	• Reviewing behavior goals • Feedback on behavior • Self-monitoring of behavior • Social support (unspecified) • Instructions on how to perform the behavior • Habit formation • Credible source • Social reward • Reduce negative emotions • Verbal persuasion about capability
Final call	• Reviewing outcome goals

Technology-Assisted Telephone Intervention

Overview

In the research group, technology was used for supporting both the coach and participants throughout the intervention. Web tools and wearables were used to support the identification of the participants' behavior change targets, the creation of the initial intervention plan, progress monitoring, and communication. Technology was designed to help coaches obtain an accurate and comprehensive picture of the participants' situation (needs and motivation) efficiently in a systematic manner to enhance coaching quality and reduce the time needed to acquire this in-depth knowledge, as well as to empower the participants to be more active in planning and performing their behavior change actions. These technological tools were new to coaches. Before using the tools, the coaches were offered training during four 6-hour information sessions, where the study protocol and other practical issues were also presented. The technology used was free for the participants, and they were encouraged to use it in their everyday lives.

At the beginning of the intensive phase, the research group received Firstbeat heart rate variability (HRV) sensors and wore them for 3 days (Firstbeat Technologies Ltd, see more details below in Firstbeat Well-being Analysis section) [62]. The coach prepared for the first coaching call by reviewing the results of the Firstbeat well-being analysis and the health recommender system's (HRS's) behavior change needs analysis (see more details below in HRS: Behavior Change Needs Analysis and Coaching Task Recommendation section). The first call was about discussing the change needs (based on Firstbeat well-being analysis and the HRS's behavior change need analysis), agreeing with a high-level behavior change goal (eg, sleep better and

manage workload), and guiding the participant to select up to 3 coaching tasks via the HRS before the next call. The participants could select tasks either from a recommended list of items or from a task library that includes all the task items available or create their own tasks. After each call, the coaches made notes related to the call (what was agreed and how long the call lasted). Before the second call, the coach reviewed the preselected tasks. During the second call, the preselected tasks were adjusted and the coaching plan was finalized (selecting goals and tasks). If physical activity tasks were selected and the participant wanted a wearable, a Fitbit activity bracelet (Fitbit Inc) [63] was provided. The selected tasks were transferred automatically to the Movendos coaching web service (more details below in Movendos Coaching Service section) [64,65]. For stress management tasks, the Oiva stress management web service was offered to the participants [58]. Coaching calls 2 to 5 were used to discuss the suitability of the goals, supporting the change through motivational interviewing, and updating the coaching plan. Before each call, the coach reviewed the progress from Movendos regarding performing the agreed coaching tasks.

During the maintenance phase, the research group received coaching only via Movendos messages. The coaches were expected to send group messages to the research group once a month and personal coaching messages every 2 weeks in addition to replying to any messages from the participants weekly. Before sending the messages, the coaches checked the progress of the participants on Movendos. The coaching messages then focused on motivating them to perform tasks that did not progress. The research group repeated the Firstbeat well-being analysis at the end of the maintenance phase, and the appropriate timings for the Firstbeat measurements and the final coaching call were agreed upon over a phone conversation between the coach and participant. The sixth and the last coaching call was used for going through the results (Firstbeat well-being analysis), the coaching experience, and forming a plan for the time after coaching. In the following section, we describe each of the technologies used in greater detail.

Firstbeat Well-being Analysis

Firstbeat well-being analysis (Firstbeat Technologies Ltd) [62] provides an analysis of the balance between stress and recovery based on HRV. The participants wore the electrodes for 3 days, and based on the HRV data, their physiological well-being was assessed.

HRS: Behavior Change Needs Analysis and Coaching Task Recommendation

During the project, a web-based HRS was developed to analyze participants' behavior change need areas and to provide personalized recommendations for suitable behavior change actions, that is, coaching tasks, based on the identified needs. The HRS evaluated several well-being–related or lifestyle-related areas (sleep sufficiency and quality, eating rhythm, balanced diet, emotional eating, physical activity, alcohol consumption, smoking, workload management, recovery from stress, anxiety, personal values, quality of relationship, and self-esteem) based on questionnaires and the results of the Firstbeat well-being analysis. As a result, the HRS provided a report summarizing for each of these areas the strength of the behavior change need (on a scale of 1 to 5) and the readiness to change the behavior categorized by the transtheoretical model's stages of change [66]. On the basis of the behavior change needs analysis, the HRS provided a list of coaching tasks recommended for inclusion in the coaching plan. In addition, the complete list of available coaching tasks can be explored via the HRS, which includes >100 tasks regarding different areas of well-being and health behaviors. Both the coach and participants could access the information provided by the HRS.

Movendos Coaching Service

The research group used the Movendos coaching web service (version 1.27; Movendos Ltd) [64,65] for (1) communicating with the coach via messages (eg, feedback), (2) progress monitoring, and (3) receiving reminders from the coach and setting reminders for themselves if they wished. The coaches had access to a wide library of tasks in the Movendos coaching service, which they could assign to participants, and the service provided them information on the participants' progress regarding the selected tasks.

Telephone Intervention

The control group received 8 coaching calls in total: 5 in the intensive phase and 3 in the maintenance phase. Before the first coaching call, the coaches reviewed the WorkOptimum questionnaire results (administered as a part of the eligibility questionnaire) [49,50]. The first call was about discussing change needs (based on the WorkOptimum index) and forming the coaching plan (selecting goals and tasks). After each call, the coaches made notes related to the call (what was agreed and how long the call lasted). Calls 2 to 7 were used for discussing the suitability of the tasks, supporting the change through motivational interviewing, and updating the coaching plan. The eighth and final call was used for going through the repeated WorkOptimum questionnaire results, the coaching experience, and forming a plan for the time after coaching.

Outcome Measures

Overview

The pilot outcome was intervention feasibility measured primarily by the participants' self-assessed mental well-being and the total time use of the coaches for the complete coaching period and secondarily by participants' adherence to and satisfaction with coaching.

The feasibility criteria are formulated in the following manner:

- *Stop—main study not feasible* if the research group has lower well-being than the control group.
- *Continue, but modify intervention—feasible with modifications* if the research group has in comparison with the control group (1) similar well-being and similar or increased time use of coaches, (2) similar well-being and decreased time use of coaches but poorer adherence or satisfaction, or (3) improved well-being and increased time use of coaches.
- *Continue without modifications—feasible but monitor closely* if the research group has in comparison with the control group (1) improved well-being and similar or

decreased time use of coaches or (2) similar well-being, decreased time use of coaches, and similar or better adherence or satisfaction.

- *Continue without modifications—feasible as is* if the research group has improved well-being and decreased time use of coaches in combination with similar or better adherence and satisfaction.

Mental Well-being

Mental well-being was assessed using the WorkOptimum index, which is a measure of occupational health, and aims to detect work-related cognitive decline and decrease in mental well-being before developing mental health problems (Multimedia Appendix 1 [49,50]). The scoring of the index is divided into 4 categories with the following interpretations: exhaustion (score −4 or less), high-risk (score −3.9 to −2.5), at risk (score −2.4 to −1.0), and good (score −0.9 to 0.0). The WorkOptimum questionnaire evaluates the ability to recover from work and perceived mental and physical resources as well as the perceived energy level and workload. The electronic questionnaire was administered to the participants at baseline (month 0) and at the end of the intensive (month 4) and maintenance (month 9) phases.

Time Use of Coaches

The total time use of coaches was tracked during the entire intervention regarding (1) *preparation time* for the coaching calls, (2) *duration* of the coaching calls (ie, 6 calls for the research group and 8 calls for the control group), and (3) the time spent on *writing personal coaching messages* to the research group. For each coaching call, the coaches were asked to manually record the duration of the call and the time spent preparing for it (in minutes). On the basis of these results, the total time spent on coaching calls per participant was computed. The time spent on messaging had to be estimated, as the coaches considered it too laborious to record the time used for messages. The estimate was computed based on the total number of messages sent to each of the research group participants (recorded by the Movendos system log) and the results of previous studies. For the intensive phase, 5 minutes was considered as the time estimate for responding to a participant's message, which is comparable with the time it took physicians to address emails in the study by Leong et al [67]. For the maintenance phase, 10 minutes were assigned as the time estimate per message [68], as coaches were required to initiate motivating coaching messages in addition to just responding to participants. The total time used by coaches was computed by adding the total time spent on coaching calls and messages together, per participant. In addition, the mean preparation time and mean duration per coaching call were computed for each participant.

Only participants for whom all the planned coaching calls were realized and the time-keeping records were complete for both call-preparation time and call duration were included in the analysis of the total time use of coaches. For 14 participants (7 participants from both groups), the time records for the coaching calls were incomplete because one of the coaches recorded only the time spent on preparation activities but missed recording the duration of the calls. Hence, complete data were available only for 11 participants per group.

Adherence

Adherence was assessed by the dropout attrition, describing how many participants quit the intervention, and by the use adherence (inversely nonuse attrition). The use adherence comprised (1) the proportion of realized coaching calls, (2) the frequency of performing the selected coaching tasks, and (3) diligence in performing the tasks. During the intensive phase, the coaches evaluated the task performance adherence (frequency and diligence) 3 times for the research group (during coaching calls 3-5) and 4 times for the control group (during calls 2-5) via a structured interview (Multimedia Appendix 2). For each coaching task, the following 3 items were assessed: “the client performed the task less frequently than agreed,” “the client performed the task more frequently than agreed,” and “the client performed the task with diligence,” with a 5-point Likert scale (1=“strongly disagree”; 5=“strongly agree”) also having the option “I don’t know.” For the statistical analyses, the answers to the items related to task performance frequency (2 items) were combined and transformed to a scale from 1 to 9 with the following meaning: 1=“the task was performed less frequently than agreed,” 5=“the task was performed as agreed,” and 9=“the task was performed more frequently than agreed.” During the maintenance phase, the participants self-assessed their task performance adherence for each coaching task via 3 electronic questionnaires administered at months 5, 7, and 9 (after calls 6-8 for the control group; Multimedia Appendix 2). The task performance frequency was assessed with the item “how actively did you perform the coaching task?” with a 5-point Likert scale (1=“less frequently than agreed,” 3=“as agreed,” 5=“more frequently than agreed”). The task performance diligence was assessed with the item “I performed the task with diligence” with a 5-point Likert scale (1=“strongly disagree”; 5=“strongly agree”). Both items included the response option “I don’t know.” For each assessment point, the evaluations for the 2 best-evaluated tasks were considered in the analyses, as the number of selected tasks varied per person and per coaching phase. Group-level medians were calculated for the intensive and maintenance phases over all the tasks considered.

Satisfaction With Coaching

Participants’ satisfaction with coaching was assessed using 1 question in different phases of the trial. For the research group, during the intensive phase, the statement was “I was satisfied with the coaching call,” and during the maintenance phase, the statement was “I was satisfied with the coaching received via Movendos messages.” For the control group, the statement remained the same throughout the intervention, that is, “I was satisfied with the coaching call.” The 5-point Likert scale was used (1 = “Strongly disagree”; 5 = “Strongly agree”). Group-level medians were calculated over all the assessments available for each group for the intensive and maintenance phases, as well as for the entire intervention. Satisfaction was assessed at 4 time points (after calls 2-5 for the research group, and after calls 3-5 for the control group) during the intensive

phase and 4 time points in the maintenance phase (after calls 6-8 for the control group).

Statistical Methods

For the primary trial outcome (WorkOptimum index for mental well-being), Mann-Whitney *U* tests were conducted on the follow-up scores measured at the end of the intensive and maintenance phases to determine the statistical significance of the between-group differences. Analysis of covariance was considered for the statistical tests, but as the residuals were not normally distributed, it was more appropriate to use nonparametric tests instead of parametric ones. Similar between-group analyses were performed for the time use of coaches and participants' adherence to (frequency and diligence) and satisfaction with coaching. In addition, the statistical significance of the within-group changes in mental well-being from baseline (month 0) to the end of the intensive (month 4) and maintenance phases (month 9) were determined using the Sign test. For both groups, the number of participants with positive changes (N_+) was reported as a test statistic for the Sign test. For the different outcomes, all participants with relevant data available were included in the analyses (available-case analysis). A significance level (α) of .05 was used for the statistical tests. Statistical tests were performed using IBM SPSS Statistics (version 25).

The Vargha-Delaney A measure of stochastic superiority [69] was reported as an indicator of the effect size. For the between-group analyses, 95% CIs of the effect sizes were reported. The effect size computations were performed with free R (version 4.0.5) statistical software by using the *rcompanion* package. The 95% CIs were computed using the bootstrap procedure [70].

Ethics Approval

This study was approved by the Ethics Committee of Human Sciences at the University of Oulu. The RCT was registered at ClinicalTrials.gov (NCT02445950). Informed consent was obtained from interested individuals through regular mail before administering the electronic eligibility survey via email.

Results

Participants

In total, 131 volunteers registered for the study, of which 56 (42.7%) met the inclusion criteria and were randomized equally to the research and control groups. Of the 56 randomized group of participants, 50 (89%) were chosen to be enrolled in the study based on the order of registration. The remaining 6 participants were put on a waiting list in case of last-minute changes in participation before starting the coaching program. At the beginning of the coaching program, 1 participant in the research group was no longer eligible for the study because of a change in their employment status and was therefore omitted from the statistical analyses. Figure 2 summarizes the participant flow from registration to available-case analysis for mental well-being (primary outcome) and the attrition numbers for the intensive and maintenance phases together with the reasons for withdrawal.

The baseline characteristics of the study participants are presented in Table 3. Most participants (47/49, 96%) were female, and the 2 (4%) male participants were allocated to the control group. More than half (28/49, 57%) of the participants were aged between 46 and 60 years (mean age 46.26, SD 9.74 years). A majority had at least a bachelor's degree (41/49, 84%). More than half (28/49, 57%) of the participants had a lower socioeconomic status, and approximately half (25/49, 51%) of the participants had children (school-aged or younger). The 2 groups had similar characteristics apart from education level, which was higher in the control group (4/24, 17%, vs 10/25, 40% having a graduate or doctoral degree). Mental well-being at the baseline was at the "at risk" level for both groups (−2.38 vs −2.14) in terms of the interpretation of the median WorkOptimum index (primary outcome).

Figure 2. Participant flow for the primary analysis regarding mental well-being.

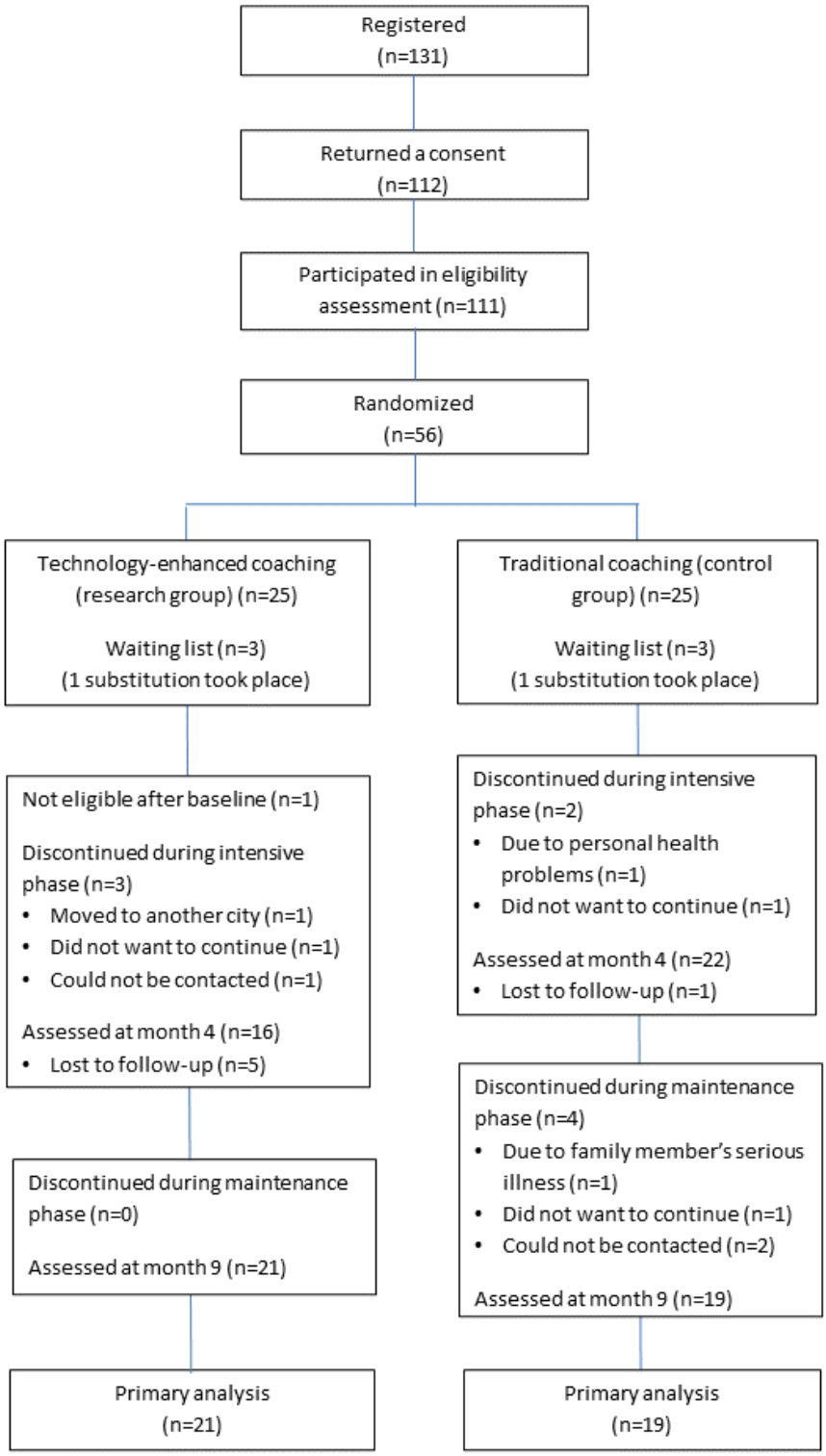


Table 3. Baseline characteristics.

	Research (n=24), n (%)	Control (n=25), n (%)	All (n=49), n (%)
Gender			
Female	24 (100)	23 (92)	47 (96)
Age (years)			
26-35	5 (21)	4 (16)	9 (18)
36-45	6 (25)	6 (24)	12 (24)
46-60	13 (54)	15 (60)	28 (57)
Education			
Secondary school	5 (21)	3 (12)	8 (16)
Bachelor's degree	15 (63)	12 (48)	27 (55)
Graduate or doctoral degree	4 (17)	10 (40)	14 (29)
Socioeconomic status[a]			
Lower-level employees	13 (54)	15 (60)	28 (57)
Upper-level employees	11 (46)	10 (40)	21 (43)
Family			
No children	13 (54)	11 (44)	24 (49)
At least 1 child below school age (<7 years)	4 (17)	5 (20)	9 (18)
Only school-aged children	7 (29)	9 (36)	16 (33)

[a]Socioeconomic groups are based on the classification of Statistics Finland [55]. Lower-level employees include, for example, secretaries, nurses, childminders, customer service staff, and assistants. Upper-level employees include, for example, managers, doctors, psychologists, teachers, and information technology professionals.

Mental Well-being

The follow-up scores for the mental well-being measure, WorkOptimum index (primary outcome), at the end of intensive (month 4: −0.95 vs −1.09, Â=0.53, 95% CI 0.34-0.72; *P*=.74) and maintenance (month 9: −0.47 vs −0.44, Â=0.56, 95% CI 0.37-0.74; *P*=.56) phases were not statistically significantly different between the research and control groups (Table 4). At month 9, the WorkOptimum index had improved from the baseline's "at risk" level (−2.4 to −1.0) to a "good" (−0.9 to 0.0) level in both groups. The within-group improvements were strong for both groups (Â>0.85, *P*≤.001 at month 9; Table 5). The distribution of WorkOptimum value categories across time points is presented in Table 6.

Table 4. Between-group differences of the outcome measures.

Outcome	Research		Control		Mann-Whitney *U* statistic	*P* value	Â[a] (95% CI)
	Median (25th; 75th) or % (range)	n	Median (25th; 75th) or % (range)	n			
Mental well-being[b]							
Month 0	−2.38 (−3.09; −1.19)	24	−2.14 (−5.50; −1.28)	25	N/A[c]	N/A	N/A
Month 4	−0.95 (−1.33; −0.50)	16	−1.09 (−1.80; −0.46)	22	188.00	.74	0.53 (0.34-0.72)
Month 9	−0.47 (−0.72; −0.15)	21	−0.44 (−1.73; −0.26)	19	221.50	.56	0.56 (0.37-0.74)
Time use of the coaches (minutes)							
Total time use							
Months 0-9	366.0 (320.0; 427.0)	11	343.0 (268.0; 489.0)	11	49.0	.48	0.60 (0.33-0.85)
Mean preparation time per call							
Months 0-9	30.00 (24.0; 32.60)	11	17.57 (16.14; 23.0)	11	12.50	.001	0.90 (0.75-1.0)
Mean duration per call							
Months 0-9	25.50 (20.33; 30.50)	17	22.44 (17.28; 32.44)	14	97.0	.40	0.59 (0.37-0.80)
Adherence							
Dropout attrition (%)							
Months 0-4	12.5[d]	24	8[d]	25	N/A	N/A	N/A
Months 4-9	0[d]	24	16[d]	25	N/A	N/A	N/A
Months 0-9	13[d]	24	24[d]	25	N/A	N/A	N/A
Mean adherence to coaching calls (%)							
Month 9	92 (50.0-100.0)	24	86 (25.0-100.0)	25	N/A	N/A	N/A
Frequency of performing the tasks							
Month 4 (scale 1 to 9)	5.0 (2.0; 5.0)	23	5.0 (2.0; 5.0)	24	6543.50	.95	0.50 (0.42-0.57)
Month 9 (scale 1 to 5)	3.0 (2.0; 3.0)	20	3.0 (2.0; 3.0)	16	2126.0	.37	0.54 (0.45-0.62)
Diligence in performing the tasks (scale 1 to 5)							
Month 4	5.0 (4.0; 5.0)	24	4.0 (3.0; 5.0)	25	5282.50	.03	0.58 (0.51-0.65)
Month 9	4.0 (30; 5.0)	20	4.0 (3.0; 5.0)	15	1973.50	.15	0.57 (0.47-0.66)
Satisfaction with coaching (scale 1 to 5)							
Months 0-4	5.0 (4.0; 5.0)	24	4.0 (4.0; 5.0)	25	1923.50	<.001	0.66 (0.58-0.73)
Months 4-9	4.50 (3.25; 5.0)	24	5.0 (4.0; 5.0)	25	938.0	.33	0.55 (0.45-0.67)
Months 0-9	5.0 (4.0; 5.0)	24	4.0 (4.0; 5.0)	25	5729.50	.03	0.58 (0.51-0.65)

[a]Vargha-Delaney *A* measure of stochastic superiority for effect size estimation. Limits for interpretation: 0.56 (small), 0.64 (medium), 0.71 (large). Between-group differences are observed when the lower bound of the 95% CI is >0.5 [69].

[b]The scoring of the index is divided into 4 categories with the following interpretations: exhaustion (score −4 or less), high-risk (score −3.9 to −2.5), at risk (score −2.4 to −1.0), and good (score −0.9–0.0).

[c]N/A: not applicable.

[d]Range is not applicable (dropout attrition describes how many people dropped out of the study).

Table 5. Changes in well-being over time.

Outcome, group, and period	Change		N+ (Sign test)	*P* value (exact)	Â[a]
	Median (25th; 75th)	n			
Mental well-being					
Research					
Months 0-4	0.54 (0.12; 1.33)	16	13.0	.02	0.81
Months 0-9	0.98 (0.29; 2.33)	21	18.0	.001	0.86
Control					
Months 0-4	2.11 (0.43; 4.60)	22	20.0	<.001	0.91
Months 0-9	2.41 (1.01; 4.56)	19	18.0	<.001	0.95

[a]Vargha-Delaney *A* measure of stochastic superiority for effect size estimation. Limits for interpretation: 0.56 (small), 0.64 (medium), 0.71 (large) [69].

Table 6. Distribution of WorkOptimum value categories at different time points for both groups (n=49).

Time point	Research (n=24), n (%)	Control (n=25), n (%)
Month 0		
Exhaustion	5 (21)	7 (28)
High risk	6 (25)	4 (16)
At risk	9 (38)	12 (48)
Good	4 (17)	2 (8)
Missing	0 (0)	0 (0)
Month 4		
Exhaustion	2 (8)	3 (12)
High risk	1 (4)	0 (0)
At risk	5 (20)	9 (36)
Good	8 (33)	10 (40)
Missing	8 (33)	3 (12)
Month 9		
Exhaustion	2 (8)	1 (4)
High risk	0 (0)	2 (8)
At risk	2 (8)	4 (16)
Good	17 (79)	12 (48)
Missing	3 (12.5)	6 (24)

Time Use of Coaches

The total time use of coaches during the 9-month coaching period was not statistically significantly different between the 2 groups (366.0 vs 343.0 minutes, Â=0.60, 95% CI 0.33-0.85; *P*=.48; Table 4). However, the mean preparation time per coaching call was considerably higher in the research group (Â=0.90, 95% CI 0.75-1.0; *P*=.001). The mean duration per call did not differ between the groups (Â=0.59, 95% CI 0.37-0.80; *P*=.40). Regarding personal coaching messages, the coaches sent a total of 60 and 102 personal coaching messages to the research group participants during the intensive and maintenance phases, respectively. The median number of messages sent per participant was 2.0 (range 0-6) for the intensive and 4.0 (range 0-19) for the maintenance phase. Therefore, the estimated median time used for coaching messaging per participant was 10.0 minutes in the intensive phase and 40.0 minutes in the maintenance phase.

Adherence

The dropout attrition rates differed between the study groups: 13% (3/24) in the research group and 24% (6/25) in the control group.

The use adherence differed somewhat between the study groups. Owing to the higher dropout attrition in the control group, the mean proportion of realized coaching calls was lower compared with the research group (86% vs 92%). There were no between-group statistical differences in task performance frequency for either phase (Â=0.50, 95% CI 0.42-0.57; *P*=.95, and Â=0.54, 95% CI 0.45-0.63; *P*=.37). The research group

performed the tasks with slightly better diligence (Â=0.58, 95% CI 0.51-0.65; *P*=.03) during the intensive phase, but no differences were observed in the maintenance phase (Â=0.57, 95% CI 0.47-0.66; *P*=.15).

Satisfaction With Coaching

During the intensive phase (months 0-4), the research group was moderately more satisfied with the coaching program than the control group (Â=0.66, 95% CI 0.58-0.73; *P*<.001). During the maintenance phase (months 4-9), satisfaction did not differ between the groups (Â=0.55, 95% CI 0.45-0.67; *P*=.33). In general, participant satisfaction was high throughout the coaching program in both groups (Table 4).

Discussion

The aim of this study was to investigate whether technology-assisted telephone intervention is feasible for increasing well-being or decreasing the time use of coaches while maintaining participants' adherence and satisfaction compared with traditional telephone intervention in an occupational health care setting.

Principal Findings

The technology-assisted telephone intervention was similarly effective in increasing well-being as traditional telephone coaching while having better adherence in 2 of the 4 metrics (lower dropout rate and higher adherence to calls) and higher satisfaction during the intensive phase. However, technology-assisted telephone coaching was unable to demonstrate savings in the time use of coaches. Therefore, the intervention is feasible, but some modifications are needed before moving on to a large-scale RCT.

The similar well-being improvements in both groups might be due to the similarities in coaching content and having human contact in both groups. In addition, it might be due to the similar baseline well-being state being only at the "at risk" level for both groups. If a person is in a poorer condition, it might be more difficult to improve mental well-being with technology or less intensive coaching, but this remains to be studied among people with poorer well-being. It is noteworthy that for the technology-assisted group, the improvement in well-being holds, although the coaching calls were replaced with personal coaching messages during the maintenance phase. Therefore, it seems feasible to replace at least some of the coaching calls with messages.

Although the total time use of coaches was similar for the 2 groups, the coaches spent more time preparing for the coaching calls of the technology-assisted group participants. The potential reason for this could be that the used technology components generated additional information on the participant's situation, which the coaches had to review before the coaching calls (Firstbeat well-being analysis, HRS's behavior change needs analysis and coaching task recommendations, and participants' progress in the Movendos Coaching Platform). However, the use of technology for providing a comprehensive picture of participants' situations (needs, motivation, and progress) should be considered an advantage, as this supports coaches in making better decisions when identifying suitable behavior change goals and activities for the coaches; this type of information could also help coaches gain a better understanding of coaches' motivation levels and the appropriate means to motivate them. Therefore, using analytical technological tools could lead to improved coaching quality and intervention effectiveness. This line of thinking is also supported by the observation that participants in the technology-assisted group reported higher satisfaction with coaching during the intensive phase than did the traditional intervention group. In addition, the used technology and technology-assisted coaching process were new to the coaches. Learning to use the new tools effectively as part of their coaching process must have taken extra time, while following the familiar telephone coaching process was obviously more efficient. There will be potential to enhance the use efficiency of new technology as they become familiar to end users.

The technology-assisted group was more persistent at staying in the intervention until the end and adhering to the coaching calls considering the dropout attrition rate was higher in the fully human intervention group. This might be due to the generally less effort needed from the technology-assisted group participants because there were no coaching calls in the maintenance phase. Perhaps it is easier to adhere to the lower number of calls. However, after the intensive phase, the technology-assisted group had a higher dropout rate than the traditional intervention group, but this difference was due to only 1 person. Moving to another town was the reason for 1 of 3 dropouts. The satisfaction for the 2 other dropouts was good and, therefore, does not explain the dropout. The task performance frequency was similar between the 2 groups. Better diligence in the technology-assisted group in the intensive phase could have been because of extensive analysis of behavior change needs at the beginning, which might have increased the personal understanding of why the tasks were important, thus leading to higher levels of diligence in performing them.

The higher levels of satisfaction for the technology-assisted group in the intensive phase might be explained by the technology providing the coaches with a deeper understanding of participants' situations and needs, which in turn may have facilitated coaches to provide the right kind of support. From this perspective, technology was fulfilling one of its goals in creating a comprehensive picture of the participant and enhancing coaching quality. However, satisfaction did not differ between the 2 groups during the maintenance phase, which is a good result because the coaching took place via messaging instead of phone calls for the technology-assisted group during this phase.

It is also interesting that 86% (42/49) of participants chose a task related to physical activity as part of the coaching plan. There are several potential reasons for this finding. One reason could be that many of the participants were employed by the health care sector and, hence, had knowledge of the basic principles of a healthy lifestyle. Coaches noted that participants commonly felt that they did not have enough time for physical activity. In Finnish culture, it is typical to think that exercise is a medicine for almost anything. Furthermore, the Fitbit wrist device offered to the technology-assisted group might have

encouraged the focus on physical activity, as it enabled easy and motivating self-monitoring of activity.

Comparison With Previous Studies

Our findings strengthen the evidence for technology-assisted human interventions being equally effective as traditional human interventions. However, the majority of the studies compared interventions with a different level of human-technology involvement from ours, where human coaching was partly or entirely replaced by technology.

Only 1 study considered the effectiveness of an intervention similar to this study, but it was compared with a fully technological intervention [23]. In this study, the technology-assisted intervention supported the coaches by providing comprehensive knowledge regarding the participant's situation and was used as communication media, but it did not provide coaching without the coach.

Our findings could not confirm earlier claims of saving time with technology use [20]. However, earlier approaches were also different from this study in the way that technology usually replaced or automated coaching tasks. This study, however, provides more information on the coaches' time use for the preparation of coaching calls and the actual call durations, which helps target changes to the intervention so that it might save time.

Adherence in both groups in terms of dropout rate (13% and 24%) was good. The adherence to earlier well-being interventions varied significantly. In another technology-assisted human (physical activity) intervention study, adherence to the intervention was 28.4% (dropout rate was 71.6%) [23]. In human-assisted technological stress management interventions, adherence varied between 69.7% and 96% [31,71,72] and, in fully technological interventions, it varied between 25% and 90.6% [23,30,31]. Thus, our findings indicate a high level of adherence.

This study provides new knowledge on satisfaction with stress management technology-assisted human interventions, showing that interventions with less intensive human involvement can lead to similar or even better participant satisfaction than more intensive human involvement.

Limitations

There were several limitations to this study. One limitation is the narrow approximation of costs based on the time use of coaches. For a comprehensive cost-effectiveness study, costs should be studied more widely from both the care provider and societal perspectives. In addition, the technology itself incurs costs that have to be considered.

Considering the study was not blinded to the participants and they knew the intervention of interest and the intervention they were participating in, it is possible that participants may have sought additional help, for example, from occupational health care to which all participants had access. Providing sufficient information is the result of balancing a valid research setting and ethics. Moreover, participants in the traditional telephone intervention group were more highly educated, which could have a positive impact on their motivation and ability to seek additional help. These issues may have affected our results, perhaps by improving the well-being of the control group.

As the used technology components and technology-assisted coaching process were new to the coaches, the study might not provide realistic results on the time use of coaches. Our study also showed that there can be many problems with human-reported metrics. There can be challenges in advising procedures, which in our case caused a misunderstanding between researchers and coaches regarding the proper procedures to record the time used for coaching. In addition, there is a risk of measurement errors when relying on self-reporting instead of objective measurements. In addition, the data collection metrics for adherence had to be improved during the pilot because the measurement scales for evaluating adherence during the intensive phase turned out to be ambiguous. The recording of time used was also stated by the coaches to be too laborious. Future studies should ensure that time logging is easy, preferably even automatic, and that it is performed consistently throughout the study.

The sample size was small, which makes the results only preliminary and must be confirmed in a larger RCT. With a small number of participants, unexpected events during the study and participant-specific variations in the selected intervention areas may have influenced the study results. The sample was also quite biased, containing mostly middle-aged, highly educated women because of the recruitment methodology. The participants were recruited from the municipal sector (employees of the City of Oulu, Finland), where especially in education, health care and social sectors, highly educated women are in majority, so it is obvious that these women are also well represented in this study. In addition, women are generally more interested in their health and more eager to participate in studies [73]. For better generalizability of the results, it would be essential to recruit participants from a group that has more variety in terms of age, gender, and education level. This could be achieved using several occupational care providers that have customers from different industries (eg, also including the manufacturing industry, where most of the employees are men).

The results were obtained in the coaching environment in which phone calls were the primary means of coaching and may not be generalized to other forms of coaching. In addition, coaching can be implemented in various ways, which makes proper comparison of different interventions or studies difficult. Individual situations and health statuses vary considerably between people, and here the participants had moderate baseline well-being. Therefore, it is difficult to generalize any results from this study, and the results only hold for this particular sample, set of tools, and processes. In addition, in the absence of a no-treatment group, the role of other factors, such as the research setting itself, cannot be quantified.

Implications

The feasibility of the technology-assisted intervention was compromised because of the inability to show a decrease in the time use of coaches. This may be due to an inefficient coaching process or unfamiliarity with the used technology, subjective

(error-prone) data collection methods, and the small sample size.

It is expected that the process can save coaches' time once the technology and its optimal use practices have been honed and become more of a routine. Therefore, coaching technology should be used for some time before starting a larger RCT. Studies should also explore which coaching activities could be further automated to maintain the effects but decrease the time use of coaches. During action planning, it is important to provide concise reports that are easy and quick for coaches to read and understand. In addition, using templates for messaging can be helpful in decreasing the time use of coaches.

There is a need for reliable objective data collection methods for the time use of coaches. It would be ideal if the data could be collected automatically, and with digital interventions, this becomes possible. Technology can support research data collection by prompting the coach after each coaching event to note how much time was used for preparation and by automatically recording the coaching call durations and writing messages, for example.

The effect size and adherence rates allowed for an estimation of the number of participants to be enrolled in the fully powered RCT. A small effect size (Â=0.56) was obtained for the difference in well-being between the groups in the end. Thus, it would be required to enroll 125 (58+67) participants (considering 13%-24% dropouts; power=0.80; significance level, α=.05) in a fully powered 2-arm RCT to study the difference in effects on well-being [74].

Conclusions

The studied technology-assisted telephone intervention is feasible with some modifications by showing similar preliminary effectiveness as the traditional telephone intervention, better satisfaction, and better adherence in 2 out of 4 adherence indicators. However, because it did not reduce the time use of coaches, it requires modifications before conducting a large-scale RCT. On the one hand, these adjustments should include adding features to the technology components to support further the work of coaches, for instance, by providing the available data regarding participants' situation in a more digested format that is fast to comprehend and providing templates for personal coaching messages. On the other hand, the protocol should be improved by recruiting more participants, using objective and automatic time-tracking methods and starting the study once the coaches have established a routine of using the technology components as part of the coaching process. Technology seems promising in terms of facilitating less resource-intensive personal coaching by replacing some coaching calls with coaching messages, but further studies are required to confirm this.

Acknowledgments

This study was supported by ARTEMIS-IA (Advanced Research & Technology for Embedded Intelligent Systems Industry Association) and TEKES (currently Business Finland) under grant 332885 (WITH-ME). The authors warmly thank Juha Leppänen and Hannu Mikkola for implementing the health recommender system and its interface on the Movendos Coaching Platform. Special thanks to Tero Myllymäki (Firstbeat Technologies Ltd) and Firstbeat for providing the Firstbeat well-being analysis technology for the project and for supporting its use in the health recommender system and the intervention. The authors also thank Miikka Ermes, PhD, for his valuable guidance on the randomized controlled trial study design; Anna-Leena Orsama, PhD, for conducting the randomization for the study, providing guidance on the statistical methods, and reviewing the manuscript; Juho Merilahti, PhD, for his support in realizing this research; Mikko Lindholm, Licentiate of Science (Technology), for his analyses on the task statistics; and Elina Mattila, PhD, for reviewing the early version of the manuscript and providing helpful comments. The authors further extend their thanks to the coaches and participants.

Authors' Contributions

AMH, HON, U-MJ, and JKK contributed to the design of the randomized controlled trial study; STM, U-MJ, and AMH were involved in conducting the study; AMH and STM performed the statistical analyses and interpretation; AMH, HON, and U-MJ specified algorithms or content for the health recommender system; STM and AMH drafted the paper, and all coauthors critically revised the paper and approved the final version for publication.

Conflicts of Interest

U-MJ works at Luona Hoiva Ltd (previously Mawell Care Ltd), which provided coaching for the intervention. HON works at Movendos Ltd, which provided the web coaching platform for the intervention.

Multimedia Appendix 1
Mental well-being—WorkOptimum.
[DOCX File , 18 KB - jmir_v24i7e26569_app1.docx]

Multimedia Appendix 2
Use adherence—frequency of and diligence in performing the tasks.
[DOCX File , 23 KB - jmir_v24i7e26569_app2.docx]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 378 KB - jmir_v24i7e26569_app3.pdf]

References

1. Akerstedt T. Psychosocial stress and impaired sleep. Scand J Work Environ Health 2006 Dec;32(6):493-501 [FREE Full text] [Medline: 17173205]
2. Van Laethem M, Beckers DG, Kompier MA, Dijksterhuis A, Geurts SA. Psychosocial work characteristics and sleep quality: a systematic review of longitudinal and intervention research. Scand J Work Environ Health 2013 Nov;39(6):535-549 [FREE Full text] [doi: 10.5271/sjweh.3376] [Medline: 23939656]
3. Kivimäki M, Nyberg S, Batty G, Fransson E, Heikkilä K, Alfredsson L, IPD-Work Consortium. Job strain as a risk factor for coronary heart disease: a collaborative meta-analysis of individual participant data. Lancet 2012 Oct 27;380(9852):1491-1497 [FREE Full text] [doi: 10.1016/S0140-6736(12)60994-5] [Medline: 22981903]
4. Cohen BE, Edmondson D, Kronish IM. State of the art review: depression, stress, anxiety, and cardiovascular disease. Am J Hypertens 2015 Nov;28(11):1295-1302 [FREE Full text] [doi: 10.1093/ajh/hpv047] [Medline: 25911639]
5. Stansfeld S, Candy B. Psychosocial work environment and mental health--a meta-analytic review. Scand J Work Environ Health 2006 Dec;32(6):443-462 [FREE Full text] [doi: 10.5271/sjweh.1050] [Medline: 17173201]
6. Wu S, Zhu W, Wang Z, Wang M, Lan Y. Relationship between burnout and occupational stress among nurses in China. J Adv Nurs 2007 Aug;59(3):233-239. [doi: 10.1111/j.1365-2648.2007.04301.x] [Medline: 17590211]
7. Khalatbari J, Ghorbanshiroudi S, Firouzbakhsh M. Correlation of job stress, job satisfaction, job motivation and burnout and feeling stress. Procedia Social Behav Sci 2013 Jul;84:860-863. [doi: 10.1016/j.sbspro.2013.06.662]
8. Chandola T. Stress at Work. London, UK: The British Academy; 2010.
9. Bonde JP. Psychosocial factors at work and risk of depression: a systematic review of the epidemiological evidence. Occup Environ Med 2008 Jul;65(7):438-445. [doi: 10.1136/oem.2007.038430] [Medline: 18417557]
10. Madsen IE, Nyberg ST, Magnusson Hanson LL, Ferrie JE, Ahola K, Alfredsson L, IPD-Work Consortium. Job strain as a risk factor for clinical depression: systematic review and meta-analysis with additional individual participant data. Psychol Med 2017 Jun;47(8):1342-1356 [FREE Full text] [doi: 10.1017/S003329171600355X] [Medline: 28122650]
11. Bourbonnais R, Mondor M. Job strain and sickness absence among nurses in the province of Québec. Am J Ind Med 2001 Feb;39(2):194-202. [doi: 10.1002/1097-0274(200102)39:2<194::aid-ajim1006>3.0.co;2-k] [Medline: 11170161]
12. Moreau M, Valente F, Mak R, Pelfrene E, de Smet P, De Backer G, et al. Occupational stress and incidence of sick leave in the Belgian workforce: the Belstress study. J Epidemiol Community Health 2004 Jun;58(6):507-516 [FREE Full text] [doi: 10.1136/jech.2003.007518] [Medline: 15143121]
13. Stress and Psychological Disorders in Great Britain 2013. Health and Safety Executive. 2013. URL: https://static.guim.co.uk/ni/1412687972727/stress-1.pdf [accessed 2022-06-22]
14. Van den Broek K, Hassard J, Flemming D, Gründler R, Dewe P, Teoh K, et al. Calculating the costs of work-related stress and psychosocial risks : literature review. European Agency for Safety and Health at Work 2015 [FREE Full text] [doi: 10.2802/20493]
15. Götz S, Hoven H, Müller A, Dragano N, Wahrendorf M. Age differences in the association between stressful work and sickness absence among full-time employed workers: evidence from the German socio-economic panel. Int Arch Occup Environ Health 2018 May;91(4):479-496 [FREE Full text] [doi: 10.1007/s00420-018-1298-3] [Medline: 29487994]
16. Brborovic H, Brborovic O, Mustajbegovic J. Looking for the possible association between stress, presenteeism and absenteeism among Croatian nurses: a cross-sectional study. Iran J Psychiatry Behav Sci 2016 Oct 24;10(4):e4587. [doi: 10.17795/ijpbs-4587]
17. Cooper CL, Lu L. Presenteeism at Work. Cambridge, UK: Cambridge University Press; 2018.
18. Schmidt B, Schneider M, Seeger P, van Vianen A, Loerbroks A, Herr R. A comparison of job stress models: associations with employee well-being, absenteeism, presenteeism, and resulting costs. J Occup Environ Med 2019 Jul;61(7):535-544. [doi: 10.1097/JOM.0000000000001582] [Medline: 30870396]
19. Mariné A, Seyrl B, Samele C, Breucker G, Miniska I, Scheftlein J, et al. Economic analysis of workplace mental health promotion and mental disorder prevention programm. Matrix Insight. 2012. URL: http://www.mentalhealthpromotion.net/resources/matrix_2012-economic-analysis-of-wmhp-programmes.pdf [accessed 2022-06-22]
20. Erbe D, Eichert H, Riper H, Ebert D. Blending face-to-face and Internet-based interventions for the treatment of mental disorders in adults: systematic review. J Med Internet Res 2017 Sep 15;19(9):e306 [FREE Full text] [doi: 10.2196/jmir.6588] [Medline: 28916506]
21. Jonas B, Leuschner F, Tossmann P. Efficacy of an internet-based intervention for burnout: a randomized controlled trial in the German working population. Anxiety Stress Coping 2017 Mar;30(2):133-144. [doi: 10.1080/10615806.2016.1233324] [Medline: 27602992]
22. Ebert D, Heber E, Berking M, Riper H, Cuijpers P, Funk B, et al. Self-guided Internet-based and mobile-based stress management for employees: results of a randomised controlled trial. Occup Environ Med 2016 May;73(5):315-323. [doi: 10.1136/oemed-2015-103269] [Medline: 26884049]

23. Prochaska JO, Evers KE, Castle PH, Johnson JL, Prochaska JM, Rula EY, et al. Enhancing multiple domains of well-being by decreasing multiple health risk behaviors: a randomized clinical trial. Popul Health Manag 2012 Oct;15(5):276-286. [doi: 10.1089/pop.2011.0060] [Medline: 22352379]
24. Sethi S, Campbell AJ, Ellis LA. The use of computerized self-help packages to treat adolescent depression and anxiety. J Technol Human Service 2010 Aug 31;28(3):144-160. [doi: 10.1080/15228835.2010.508317]
25. Ebert DD, Van Daele T, Nordgreen T, Karekla M, Compare A, Zarbo C, et al. Internet- and mobile-based psychological interventions: applications, efficacy, and potential for improving mental health. Eur Psychol 2018 May;23(2):167-187. [doi: 10.1027/1016-9040/a000318]
26. Zetterqvist K, Maanmies J, Ström L, Andersson G. Randomized controlled trial of Internet-based stress management. Cogn Behav Ther 2003;32(3):151-160. [doi: 10.1080/16506070302316] [Medline: 16291546]
27. Proudfoot J, Clarke J, Birch M, Whitton AE, Parker G, Manicavasagar V, et al. Impact of a mobile phone and Web program on symptom and functional outcomes for people with mild-to-moderate depression, anxiety and stress: a randomised controlled trial. BMC Psychiatry 2013 Nov 18;13:312 [FREE Full text] [doi: 10.1186/1471-244X-13-312] [Medline: 24237617]
28. Drozd F, Raeder S, Kraft P, Bjørkli CA. Multilevel growth curve analyses of treatment effects of a Web-based intervention for stress reduction: randomized controlled trial. J Med Internet Res 2013 Apr 22;15(4):e84 [FREE Full text] [doi: 10.2196/jmir.2570] [Medline: 23607962]
29. Wiegand B, Luedtke K, Friscia D, Nair M, Aleles M, McCloskey R. Efficacy of a comprehensive program for reducing stress in women: a prospective, randomized trial. Curr Med Res Opin 2010 Apr;26(4):991-1002. [doi: 10.1185/03007991003688193] [Medline: 20201624]
30. Fuller-Tyszkiewicz M, Richardson B, Little K, Teague S, Hartley-Clark L, Capic T, et al. Efficacy of a smartphone app intervention for reducing caregiver stress: randomized controlled trial. JMIR Ment Health 2020 Jul 24;7(7):e17541 [FREE Full text] [doi: 10.2196/17541] [Medline: 32706716]
31. Rose RD, Buckey JC, Zbozinek TD, Motivala SJ, Glenn DE, Cartreine JA, et al. A randomized controlled trial of a self-guided, multimedia, stress management and resilience training program. Behav Res Ther 2013 Feb;51(2):106-112. [doi: 10.1016/j.brat.2012.11.003] [Medline: 23262118]
32. Heber E, Ebert D, Lehr D, Cuijpers P, Berking M, Nobis S, et al. The benefit of Web- and computer-based interventions for stress: a systematic review and meta-analysis. J Med Internet Res 2017 Feb 17;19(2):e32 [FREE Full text] [doi: 10.2196/jmir.5774] [Medline: 28213341]
33. Wentzel J, van der Vaart R, Bohlmeijer E, van Gemert-Pijnen JE. Mixing online and face-to-face therapy: how to benefit from blended care in mental health care. JMIR Ment Health 2016 Feb 09;3(1):e9 [FREE Full text] [doi: 10.2196/mental.4534] [Medline: 26860537]
34. Karhula T, Vuorinen A, Rääpysjärvi K, Pakanen M, Itkonen P, Tepponen M, et al. Telemonitoring and mobile phone-based health coaching among Finnish diabetic and heart disease patients: randomized controlled trial. J Med Internet Res 2015 Jun 17;17(6):e153 [FREE Full text] [doi: 10.2196/jmir.4059] [Medline: 26084979]
35. Ebert DD, Kählke F, Buntrock C, Berking M, Smit F, Heber E, et al. A health economic outcome evaluation of an Internet-based mobile-supported stress management intervention for employees. Scand J Work Environ Health 2018 Mar 01;44(2):171-182 [FREE Full text] [doi: 10.5271/sjweh.3691] [Medline: 29144535]
36. Kählke F, Buntrock C, Smit F, Berking M, Lehr D, Heber E, et al. Economic evaluation of an Internet-based stress management intervention alongside a randomized controlled trial. JMIR Ment Health 2019 May 15;6(5):e10866 [FREE Full text] [doi: 10.2196/10866] [Medline: 31094355]
37. Lindsäter E, Axelsson E, Salomonsson S, Santoft F, Ljótsson B, Åkerstedt T, et al. Cost-effectiveness of therapist-guided Internet-based cognitive behavioral therapy for stress-related disorders: secondary analysis of a randomized controlled trial. J Med Internet Res 2019 Sep 13;21(9):e14675 [FREE Full text] [doi: 10.2196/14675] [Medline: 31586370]
38. Kaldo V, Levin S, Widarsson J, Buhrman M, Larsen H, Andersson G. Internet versus group cognitive-behavioral treatment of distress associated with tinnitus: a randomized controlled trial. Behav Ther 2008 Dec;39(4):348-359. [doi: 10.1016/j.beth.2007.10.003] [Medline: 19027431]
39. Buntrock C, Ebert DD, Lehr D, Cuijpers P, Riper H, Smit F, et al. Evaluating the efficacy and cost-effectiveness of Web-based indicated prevention of major depression: design of a randomised controlled trial. BMC Psychiatry 2014 Jan 31;14(1):25. [doi: 10.1186/1471-244x-14-25]
40. Kooistra L, Wiersma J, Ruwaard J, Neijenhuijs K, Lokkerbol J, van Oppen P, et al. Cost and effectiveness of blended versus standard cognitive behavioral therapy for outpatients with depression in routine specialized mental health care: pilot randomized controlled trial. J Med Internet Res 2019 Oct 29;21(10):e14261 [FREE Full text] [doi: 10.2196/14261] [Medline: 31663855]
41. Hilvert-Bruce Z, Rossouw PJ, Wong N, Sunderland M, Andrews G. Adherence as a determinant of effectiveness of Internet cognitive behavioural therapy for anxiety and depressive disorders. Behav Res Ther 2012 Aug;50(7-8):463-468. [doi: 10.1016/j.brat.2012.04.001] [Medline: 22659155]
42. 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]

43. Carolan S, Harris P, Cavanagh K. Improving employee well-being and effectiveness: systematic review and meta-analysis of Web-based psychological interventions delivered in the workplace. J Med Internet Res 2017 Jul 26;19(7):e271 [FREE Full text] [doi: 10.2196/jmir.7583] [Medline: 28747293]
44. Eisen K, Allen G, Bollash M, Pescatello L. Stress management in the workplace: a comparison of a computer-based and an in-person stress-management intervention. Comput Human Behav 2008 Mar;24(2):486-496. [doi: 10.1016/j.chb.2007.02.003]
45. Zarski A, Lehr D, Berking M, Riper H, Cuijpers P, Ebert D. Adherence to Internet-based mobile-supported stress management: a pooled analysis of individual participant data from three randomized controlled trials. J Med Internet Res 2016 Jun 29;18(6):e146 [FREE Full text] [doi: 10.2196/jmir.4493] [Medline: 27357528]
46. Ovbiosa-Akinbosoye O, Long D. Wellness program satisfaction, sustained coaching participation, and achievement of health goals. J Occup Environ Med 2012 May;54(5):592-597. [doi: 10.1097/JOM.0b013e3182496e74] [Medline: 22504959]
47. Jacobs MK, Christensen A, Snibbe JR, Dolezal-Wood S, Huber A, Polterok A. A comparison of computer-based versus traditional individual psychotherapy. Professional Psychol Res Pract 2001;32(1):92-96. [doi: 10.1037/0735-7028.32.1.92]
48. Dolezal-Wood S, Belar C, Snibbe J. A comparison of computer-assisted psychotherapy and cognitive-behavioral therapy in groups. J Clin Psychol Med Settings 1998;5:103-115. [doi: 10.1023/A:1026210020906]
49. Ahveninen H, Rintala J, Ollikainen J, Suohonen J, Arola H. Työoptimi-kysely auttaa tunnistamaan työssäkäyvän kognitiivisia ja muistioireita. Lääkärilehti 2014;4:2017-2014 [FREE Full text]
50. Ahveninen C. TyöOptimi tunnistaa tuen tarpeessa olevan työntekijän. University of Helsinki. 2016. URL: https://helda.helsinki.fi/handle/10138/163425 [accessed 2022-06-22]
51. Nordenmark M, Almén N, Vinberg S. Work/family conflict of more importance than psychosocial working conditions and family conditions for mental wellbeing. Societies 2020 Sep 14;10(3):67. [doi: 10.3390/soc10030067]
52. Miettinen A, Rotkirch A. Yhteistä aikaa etsimässä: lapsiperheiden ajankäyttö 2000-luvulla. Helsinki, Finland: Väestöliitto; 2012.
53. Modern Parenthood. Pew Research Center. 2013 Mar 14. URL: https://www.pewresearch.org/social-trends/2013/03/14/modern-parenthood-roles-of-moms-and-dads-converge-as-they-balance-work-and-family/ [accessed 2022-06-22]
54. American time use survey — 2017 results. News Release Bureau of Labor Statistics. 2018. URL: https://www.bls.gov/news.release/archives/atus_06282018.pdf [accessed 2022-06-22]
55. Classification of socio-economic groups. Statistics Finland. 1989. URL: http://www.stat.fi/meta/luokitukset/sosioekon_asema/001-1989/koko_luokitus_en.html [accessed 2019-05-21]
56. Patja K, Absetz P, Auvinen A, Tokola K, Kytö J, Oksman E, et al. Health coaching by telephony to support self-care in chronic diseases: clinical outcomes from the TERVA randomized controlled trial. BMC Health Serv Res 2012 Jun 10;12:147 [FREE Full text] [doi: 10.1186/1472-6963-12-147] [Medline: 22682298]
57. Lally P, van Jaarsveld C, Potts H, Wardle J. How are habits formed: modelling habit formation in the real world. Eur J Soc Psychol 2009 Jul 16;40(6):998-1009. [doi: 10.1002/ejsp.674]
58. Tietoisuutta arkeen. OIVA. URL: https://oivamieli.fi/ [accessed 2022-06-22]
59. Treasure J. Motivational interviewing. Adv Psychiatr Treat 2018 Jan 02;10(5):331-337. [doi: 10.1192/apt.10.5.331]
60. Prochaska JO, Redding CA, Evers KE. The transtheoretical model and stages of change. In: Glanz K, Rimer BK, Viswanath K, editors. Health Behavior: Theory, Research, and Practice. Hoboken, NJ, USA: Wiley; 2015:125-148.
61. 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]
62. Firstbeat. URL: https://www.firstbeat.com/en/ [accessed 2020-07-09]
63. Fitbit. URL: https://www.fitbit.com/ [accessed 2020-07-09]
64. Nieminen H, Baggström S, Lehtonen H, Mikkola H, Leppisaari A, Korhonen I. Mobile solution for increasing the efficiency of health coaching for preventative health care and chronic disease management. Trans Japanese Soc Med Biol Eng 2013;51(Supplement):153. [doi: 10.11239/jsmbe.51.M-153]
65. Movendos. URL: https://www.movendos.com/ [accessed 2020-07-09]
66. Prochaska J, Velicer W. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
67. Leong S, Gingrich D, Lewis P, Mauger D, George J. Enhancing doctor-patient communication using email: a pilot study. J Am Board Fam Pract 2005;18(3):180-188 [FREE Full text] [doi: 10.3122/jabfm.18.3.180] [Medline: 15879565]
68. Rosen P, Kwoh C. Patient-physician e-mail: an opportunity to transform pediatric health care delivery. Pediatrics 2007 Oct;120(4):701-706. [doi: 10.1542/peds.2007-1094] [Medline: 17908755]
69. Vargha A, Delaney H, Vargha A. A critique and improvement of the "CL" common language effect size statistics of McGraw and Wong. J Educ Behav Stat 2000 Feb;25(2):101. [doi: 10.2307/1165329]
70. Li JC. Effect size measures in a two-independent-samples case with nonnormal and nonhomogeneous data. Behav Res Methods 2016 Dec;48(4):1560-1574. [doi: 10.3758/s13428-015-0667-z] [Medline: 26487051]

71. Heber E, Lehr D, Ebert D, Berking M, Riper H. Web-based and mobile stress management intervention for employees: a randomized controlled trial. J Med Internet Res 2016 Jan 27;18(1):e21 [FREE Full text] [doi: 10.2196/jmir.5112] [Medline: 26818683]
72. Hedman E, Andersson E, Ljótsson B, Axelsson E, Lekander M. Cost effectiveness of internet-based cognitive behaviour therapy and behavioural stress management for severe health anxiety. BMJ Open 2016 Apr 25;6(4):e009327 [FREE Full text] [doi: 10.1136/bmjopen-2015-009327] [Medline: 27113231]
73. Honka A, Helander E, Pavel M, Jimison H, Mustonen P, Korhonen I, et al. Exploring associations between the self-reported values, well-being, and health behaviors of Finnish citizens: cross-sectional analysis of more than 100,000 web-survey responses. JMIR Ment Health 2019 Apr 22;6(4):e12170 [FREE Full text] [doi: 10.2196/12170] [Medline: 31008710]
74. Sample Size Calculators for designing clinical research. University of California San Francisco. URL: https://sample-size.net/sample-size-means/ [accessed 2021-05-14]

Abbreviations

HRS: health recommender system
HRV: heart rate variability
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 17.12.20; peer-reviewed by S Broetje, K Knittle, S Wen; comments to author 02.04.21; revised version received 24.06.21; accepted 19.05.22; published 13.07.22.

Please cite as:
Muuraiskangas ST, Honka AM, Junno UM, Nieminen HO, Kaartinen JK
A Technology-Assisted Telephone Intervention for Work-Related Stress Management: Pilot Randomized Controlled Trial
J Med Internet Res 2022;24(7):e26569
URL: https://www.jmir.org/2022/7/e26569
doi:10.2196/26569
PMID:35830233

Original Paper

Effects of Structured Supervised Exercise Training or Motivational Counseling on Pregnant Women's Physical Activity Level: FitMum - Randomized Controlled Trial

Signe de Place Knudsen[1,2*], MSc; Saud Abdulaziz Alomairah[1,2,3*], MPH; Caroline Borup Roland[1,2], MSc; Anne Dsane Jessen[1,2], MSc; Ida-Marie Hergel[1], MSc; Tine D Clausen[1,4], PhD; Jakob Eg Larsen[5], PhD; Gerrit van Hall[2,6], PhD; Andreas Kryger Jensen[7,8], PhD; Stig Molsted[4,8], PhD; Jane M Bendix[1,8], PhD; Ellen Løkkegaard[1,4], PhD; Bente Stallknecht[2], DMSc

[1]Department of Gynecology and Obstetrics, Copenhagen University Hospital—North Zealand, Hillerod, Denmark

[2]Department of Biomedical Sciences, University of Copenhagen, Copenhagen, Denmark

[3]Public Health Department, College of Health Sciences, Saudi Electronic University, Riyadh, Saudi Arabia

[4]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark

[5]Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kgs. Lyngby, Denmark

[6]Clinical Biochemistry, Clinical Metabolomics Core Facility, Rigshospitalet, Copenhagen, Denmark

[7]Section of Biostatistics, Department of Public Health, University of Copenhagen, Copenhagen, Denmark

[8]Department of Clinical Research, Copenhagen University Hospital—North Zealand, Hillerod, Denmark

[*]these authors contributed equally

Corresponding Author:
Signe de Place Knudsen, MSc
Department of Gynecology and Obstetrics
Copenhagen University Hospital—North Zealand
Dyrehavevej 29
Hillerod, 3400
Denmark
Phone: 45 25393022
Email: signedpi@gmail.com

Abstract

Background: Physical activity (PA) during pregnancy is an effective and safe way to improve maternal health in uncomplicated pregnancies. However, compliance with PA recommendations remains low among pregnant women.

Objective: The purpose of this study was to evaluate the effects of offering structured supervised exercise training (EXE) or motivational counseling on PA (MOT) during pregnancy on moderate-to-vigorous intensity physical activity (MVPA) level. Additionally, complementary measures of PA using the Pregnancy Physical Activity Questionnaire (PPAQ) and gold standard doubly labeled water (DLW) technique were investigated. The hypotheses were that both EXE and MOT would increase MVPA in pregnancy compared with standard care (CON) and that EXE would be more effective than MOT. In addition, the association between MVPA and the number of sessions attended was explored.

Methods: A randomized controlled trial included 220 healthy, inactive pregnant women with a median gestational age of 12.9 (IQR 9.4-13.9) weeks. A total of 219 women were randomized to CON (45/219), EXE (87/219), or MOT (87/219). The primary outcome was MVPA (minutes per week) from randomization to the 29th gestational week obtained by a wrist-worn commercial activity tracker (Vivosport, Garmin International). PA was measured by the activity tracker throughout pregnancy, PPAQ, and DLW. The primary outcome analysis was performed as an analysis of covariance model adjusting for baseline PA.

Results: The average MVPA (minutes per week) from randomization to the 29th gestational week was 33 (95% CI 18 to 47) in CON, 50 (95% CI 39 to 60) in EXE, and 40 (95% CI 30 to 51) in MOT. When adjusted for baseline MVPA, participants in EXE performed 20 (95% CI 4 to 36) minutes per week more MVPA than participants in CON (P=.02). MOT was not more effective than CON; EXE and MOT also did not differ. MVPA was positively associated with the number of exercise sessions attended in EXE from randomization to delivery (P=.04). Attendance was higher for online (due to COVID-19 restrictions) compared with physical exercise training (P=.03). Adverse events and serious adverse events did not differ between groups.

Conclusions: Offering EXE was more effective than CON to increase MVPA among pregnant women, whereas offering MOT was not. MVPA in the intervention groups did not reach the recommended level in pregnancy. Changing the intervention to online due to COVID-19 restrictions did not affect MVPA level but increased exercise participation.

Trial Registration: ClinicalTrials.gov NCT03679130; https://clinicaltrials.gov/ct2/show/NCT03679130

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2020-043671

(J Med Internet Res 2022;24(7):e37699) doi:10.2196/37699

KEYWORDS

motivation; physical activity; pregnancy; pregnant; RCT; randomized controlled trial; intervention; commercial activity tracker; tracker; COVID-19; maternal health; doubly labeled water; physical activity questionnaire; women's health; maternal; maternity; digital health; exercise; fitness; health outcome

Introduction

Physical activity (PA) is a safe and effective way to improve maternal health in uncomplicated pregnancies [1,2]. Regular PA during pregnancy reduces the risk of gestational weight gain, gestational diabetes mellitus, gestational hypertension, preeclampsia, cesarean delivery [3], and depression [4]. In addition, lifestyle interventions during pregnancy may improve offspring health by improving placental function [5,6], reducing the risk of preterm delivery [3], and normalizing birth weight [7,8]. Nevertheless, compliance with PA recommendations remains low among pregnant women worldwide [9]. Therefore, a pressing issue to address is how to implement PA in the everyday life of pregnant women.

A diverse range of approaches to PA interventions exists, of which structured supervised exercise training and motivational counseling on PA are used widely in the literature [10]. Supervised exercise training with scheduled exercise sessions provides a standard approach to increase PA in pregnant women. Recognizing the needs of an individually tailored approach [11,12], motivational counseling focuses on PA behavior has also been shown to reduce the decline or even increase PA during pregnancy [13-15]. Structured supervised exercise and motivational counseling on PA have been applied separately in studies of pregnant women [16-26], but a direct comparison of the two approaches to increase PA during pregnancy has not yet been performed.

The primary objective of FitMum was to evaluate the effects of offering structured supervised exercise training (EXE) or motivational counseling on PA (MOT) compared to standard care (CON) on moderate-to-vigorous intensity PA (MVPA) in pregnant women as determined by a wrist-worn commercial activity tracker. Secondary measures of PA were obtained by the Danish version of the Pregnancy Physical Activity Questionnaire (PPAQ-DK) [27,28] and by the gold standard doubly labeled water (DLW) technique [29-31]. The hypotheses were that both EXE and MOT would increase MVPA in pregnancy compared to CON and that EXE would be more effective than MOT [32,33]. In addition, the association between MVPA and the number of sessions attended was explored.

Methods

Ethics Approval

The study was approved by the Danish National Committee on Health Research Ethics (H-18011067) and the Danish Data Protection Agency (P-2019-512) and registered at ClinicalTrials.gov (NCT03679130). The study adheres to the principles of the Helsinki declaration. Written informed consent was obtained at inclusion.

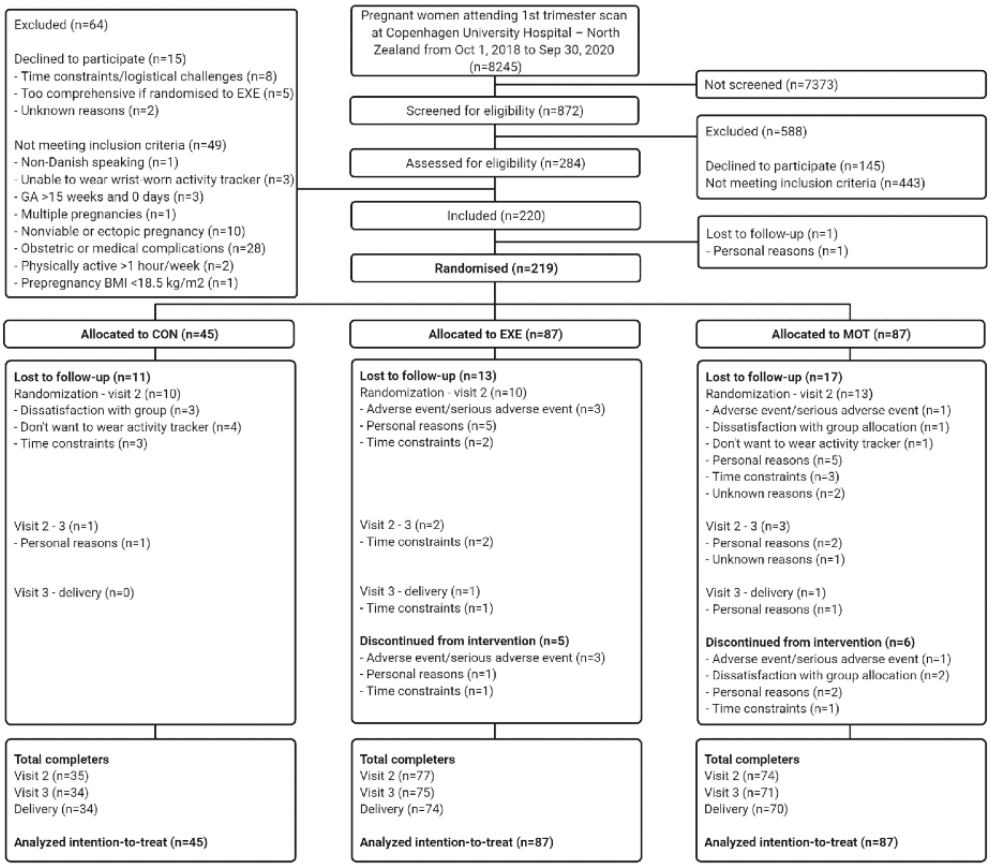
Patient and Public Involvement

The development of FitMum was inspired by stakeholders: 27 semistructured interviews with Danish pregnant women, midwives, and obstetricians were performed to explore the feasibility, facilitators, and barriers to PA during pregnancy.

Participants and Trial Design

FitMum was a single-site randomized controlled trial (RCT) conducted from 2018-2021 at the Department of Gynecology and Obstetrics at Copenhagen University Hospital–North Zealand, Denmark [32]. A total of 220 healthy, inactive pregnant women with gestational ages of ≤15 weeks and 0 days were included (visit 1). Participants were randomized 1:2:2 into CON, EXE, and MOT groups, respectively (Figure 1). Participants were invited to a test visit at the 29th gestational week (visit 2) and the 35th gestational week (visit 3).

Figure 1. Flowchart of the FitMum randomized controlled trial including enrollment, study group allocation, follow-up, and data analysis. GA: gestational age; CON: standard care; EXE: structured supervised exercise training; MOT: motivational counseling on physical activity.



Interventions

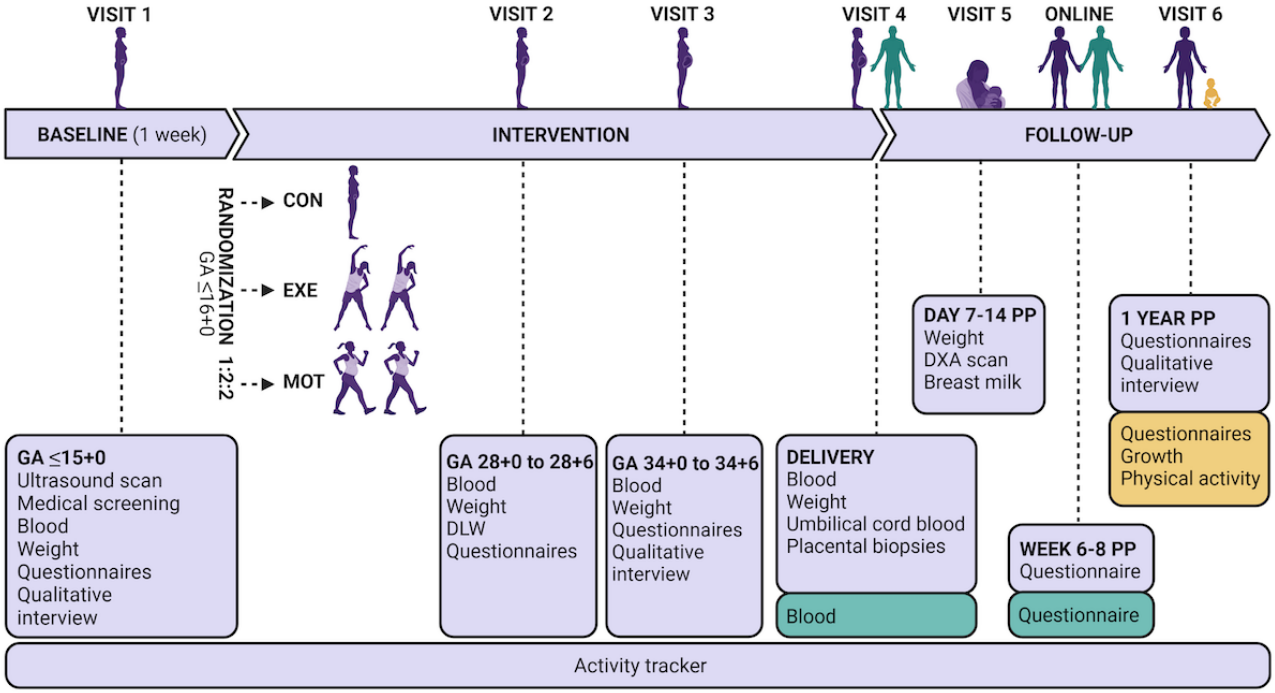
All 3 groups were offered standard maternal care. The EXE group was offered 1-hour group-based supervised exercise training at moderate intensity 3 times per week in a gym and swimming pool. The MOT group was offered 4 individual and 3 group PA motivational counseling face-to-face sessions of 1 to 2 hours duration during pregnancy and 1 weekly, personalized text message to support PA. The motivation technique applied is inspired by motivational interviewing [34], self-determination theory [35], and behavior change techniques [36].

Interventions ran from randomization until delivery. The target PA level for the EXE and MOT groups was at least 30 minutes per day at a moderate intensity as recommended in Denmark to healthy pregnant women [37]. Interventions were converted into online versions during the COVID-19 pandemic restrictions introduced in Denmark on March 11, 2020, and throughout the study period. The EXE group could access the swimming pool for 3 months during this period.

Outcome Measures

The data collection procedures are illustrated in Figure 2. PA was continuously monitored by the activity tracker from randomization to delivery, by PPAQ at visits 1, 2, and 3, and by DLW at visit 2.

Figure 2. Schedule of visits. GA: gestational age; CON: standard care; EXE: structured supervised exercise training; MOT: motivational counseling on physical activity; DLW: doubly labeled water technique; PP: postpartum; DXA: dual-energy x-ray absorptiometry.



Activity Tracker

The primary outcome was MVPA (minutes per week) from randomization to visit 2. PA was from inclusion to delivery continuously captured by a wrist-worn commercial activity tracker (Vivosport, Garmin International) [38] with a built-in heart rate monitor and accelerometer. Baseline PA was captured from inclusion to randomization (6 full days). PA with a metabolic equivalent of task (MET) value of ≥3 in bouts of at least 10 consecutive minutes was recorded automatically as MVPA by the activity tracker [38]. Secondary outcomes measured by the activity tracker were PA duration at moderate and vigorous intensities; steps; active time; active kilocalories; floors climbed; and minimum, maximum, resting, and average heart rate from randomization to delivery. At inclusion, the activity tracker was preset with PA notifications turned off and an identical face of the tracker showing only clock and battery level. After randomization, women in the MOT group were encouraged to personalize the tracker with, for example, individual goal settings and PA notifications as part of the intervention. Interaction with the tracker was neither encouraged nor controlled for the EXE and CON groups. Throughout the study period tracker software was automatically updated [38].

Danish Version of the PPAQ

PA was digitally self-reported by participants using the PPAQ-DK [28] at visits 1, 2, and 3. The questionnaire assesses PA related to everyday activities during the current trimester (eg, household, occupational, sports, and transportation) [27].

DLW Technique

Participants collected 2 baseline urine samples prior to visit 2, drank the DLW dose at the visit, and then collected and stored 5 postdose urine samples at home on days 1, 4, 7, 11, and 14 and later at −80°C. [31,39]. The calculation of total energy expenditure (TEE) was based on the Weir equation [39], and the active energy expenditure (AEE) was calculated by subtracting the basal metabolic rate (BMR) from the TEE. BMR was estimated by an equation appropriate for pregnant women [40]. PA level (PAL) was calculated by dividing TEE by BMR.

Activity Tracker Data Management

PA was transferred via Bluetooth from the activity tracker to the Garmin Connect app (Garmin International) [38] from which Fitabase (Small Steps Labs LLC) obtained the data via the programming interface. PA was monitored through Fitabase, and participants were reminded if the activity trackers were not synchronizing. PA data were downloaded from Fitabase, processed, and cleaned in R software (R Foundation for Statistical Computing).

Statistical Analyses

Statistical analyses were performed according to the statistical analysis plan, which includes a sample size calculation [33] using R. Data are presented as means and standard deviations for symmetric distributions, medians and IQRs for skewed data, and frequencies and percentages for categorical variables. The level of statistical significance was 5% except for the primary hypothesis which consisted of 2 subhypotheses; the type I error for each hypothesis test was a priori set to 2.5% to obtain a family-wise error rate of 5%. Wald-based 95% CI were given for all reported estimates [33]. Intention-to-treat analyses using all randomized participants were performed for the primary outcome. Missing observations in tracker data due to nonwear time were imputed by multiple imputations in 25 data sets using a prespecified seed, preselected baseline variables (body weight, age, PA, educational level, and parity), and the random forest imputation model from the mice R package [41]. A statistician blinded for the intervention performed the imputation and the primary outcome analysis as an analysis of covariance model

adjusting for baseline PA. MVPA before and during the COVID-19 pandemic was compared within groups with a linear regression model. Cumulative trajectories were estimated from the imputed data using a generalized additive model with a penalized regression spline with point-wise 95% confidence bands estimated by a bootstrap procedure [42]. For the PPAQ-DK outcome, a constrained linear mixed model was fitted with the observation times as a factor [43]. Both within and between-group effects were reported as estimated differences in means. For the DLW outcome, a one-way analysis of variance was used to compare the 3 group averages. For the DLW outcome, a 1-way analysis of variance was used to compare the 3 group averages. Linear regression was used to model the relationship between attended intervention sessions and attained MVPA in the EXE and MOT groups.

Results

Participants and Adherence to Interventions

In total, 220 pregnant women were included from October 2018 to October 2020. Of those, 219 were randomly allocated to CON (45/219), EXE (87/219) or MOT (87/219; Figure 1). Maternal baseline characteristics are presented in Table 1.

From randomization to visit 2, 15.1% (33/219) of participants were lost to follow-up (CON: 10/45, 22%; EXE: 10/87, 11%; MOT: 13/87, 15%). The main reason (18/33, 55%) was personal matters (eg, time consumed with participation or family reasons). From randomization to delivery, 18.7% (41/219) of participants were lost to follow-up, and proportions were similar across groups (Figure 1).

Participants randomized to EXE participated in 1.4 (95% CI 1.2 to 1.6) exercise sessions per week from randomization to visit 2, and 1.3 (95% CI 1.1 to 1.5) exercise sessions per week from randomization to delivery. Participants randomized to the MOT group joined 5.2 (95% CI 4.7 to 5.7) counseling sessions during their pregnancy.

Table 1. Baseline characteristics of randomized participants.

Characteristics	All (n=219)	CON[a] (n=45)	EXE[b] (n=87)	MOT[c] (n=87)
Age (years), mean (SD)	31.5 (4.3)	32.0 (4.6)	31.1 (4.3)	31.7 (4.1)
Gestational age at inclusion (weeks), median (IQR)	12.9 (9.4-13.9)	12.9 (9.7-13.9)	12.6 (9.3-13.7)	12.9 (9.6-13.9)
Weight (kg), mean (SD)	75.4 (15.3)	72.0 (13.7)	76.2 (17.4)	76.3 (13.8)
Prepregnancy BMI[d] (kg/m^2), median (IQR)	24.1 (21.8-28.7)	23.5 (21.3-26.8)	25.2 (21.6-29.8)	24.1 (22.4-28.9)
Nulliparity, n (%)	82 (37.4)	16 (3.56)	40 (46.0)	26 (29.9)
Educational level, n (%)				
School ≥12 years	191 (87.2)	41 (91.1)	74 (85.1)	76 (87.4)
Further education ≥3 years	175 (79.9)	33 (73.3)	73 (83.9)	69 (79.3)
Employed/studying	199 (90.9)	39 (86.7)	83 (95.4)	77 (88.5)

[a]CON: standard care.

[b]EXE: structured supervised exercise training.

[c]MOT: motivational counseling on physical activity.

[d]Prepregnancy BMI is calculated based on n=218 (CON: 45/218, EXE: 86/218, MOT: 87/218) due to a missing value.

PA by Activity Tracker

Moderate-to-Vigorous Intensity Physical Activity

The average MVPA (minutes per week) from randomization to visit 2 was 33 (95% CI 18 to 47) in CON, 50 (95% CI 39 to 60) in EXE, and 40 (95% CI 30 to 51) in MOT (Figure 3). When adjusted for baseline MVPA, participants in EXE performed 20 (95% CI 4 to 36) minutes per week more MVPA than participants in CON (*P*=.02; Multimedia Appendix 1).

The same pattern was seen throughout the entire pregnancy, hence the unadjusted average MVPA (minutes per week) was 35 (95% CI 19 to 51) in CON, 54 (95% CI 42 to 65) in EXE and 43 (95% CI 32 to 55) in MOT from randomization to delivery (Figure 3). Throughout pregnancy, participants in EXE performed 21 (95% CI 3 to 39) minutes per week more MVPA than participants in CON when adjusted for baseline MVPA (*P*=.02; Multimedia Appendix 1).

There were no significant differences in adjusted MVPA between CON and MOT (randomization to visit 2: *P*=.23; randomization to delivery: *P*=.27) or between MOT and EXE (randomization to visit 2: *P*=.14; randomization to delivery: *P*=.15; Multimedia Appendix 1).

Unplanned analysis on cumulative MVPA from randomization to delivery revealed great variability and that EXE tended to have more MVPA compared with MOT, which became significant in the late part of pregnancy (Figures 4 and 5). The same tendency was seen between CON and EXE, but the difference was insignificant. Cumulative MVPA did not differ between CON and MOT.

The number of training sessions attended in EXE from randomization to delivery was positively associated with MVPA level (P=.04). No association was present between the number of sessions attended in MOT and MVPA (P=.14).

Figure 3. Moderate-to-vigorous intensity physical activity (primary outcome) and additional activity tracker outcomes (mean and 95% CI) from randomization to visit 2 (29th week of gestation; solid line) and from randomization to delivery (dotted line). MVPA: moderate-to-vigorous intensity physical activity; CON: standard care; EXE: structured supervised exercise training; MOT: motivational counseling on physical activity.

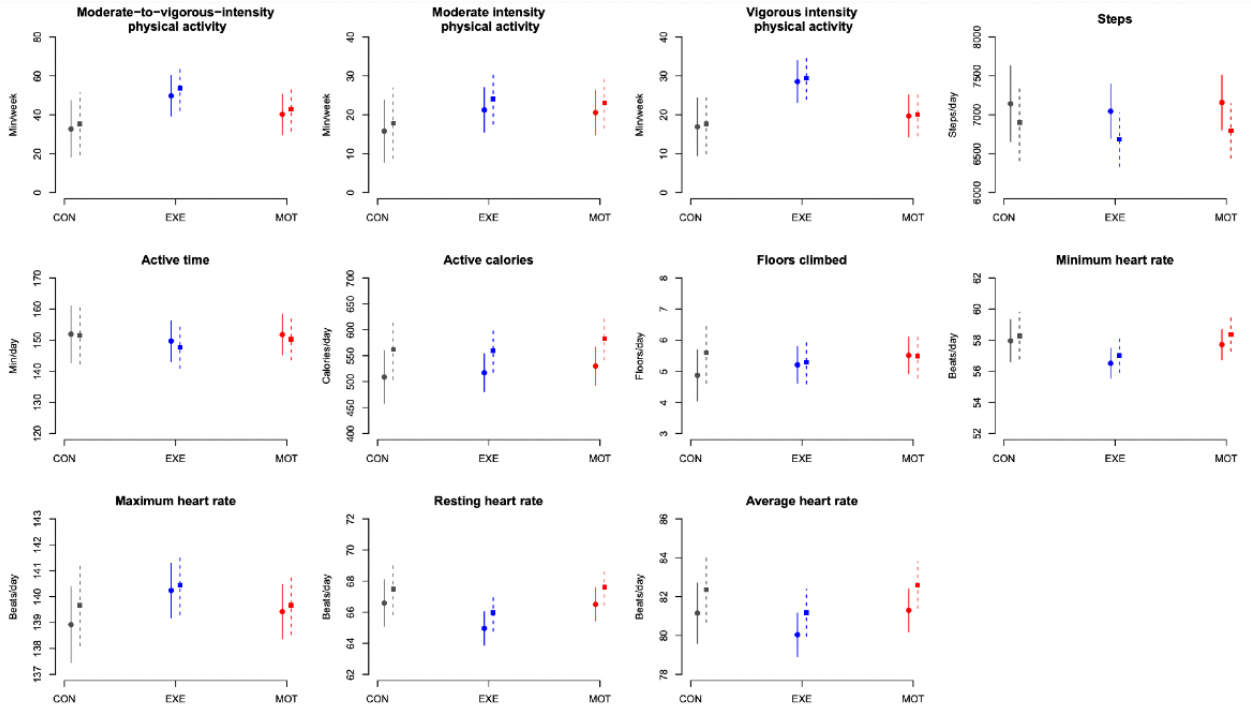


Figure 4. Cumulative moderate-to-vigorous intensity physical activity from randomization to delivery: (A) group averages, (B) EXE vs CON, (C) MOT vs CON, and (D) EXE vs MOT. MVPA: moderate-to-vigorous intensity physical activity; CON: standard care; EXE: structured supervised exercise training; MOT: motivational counseling on physical activity.

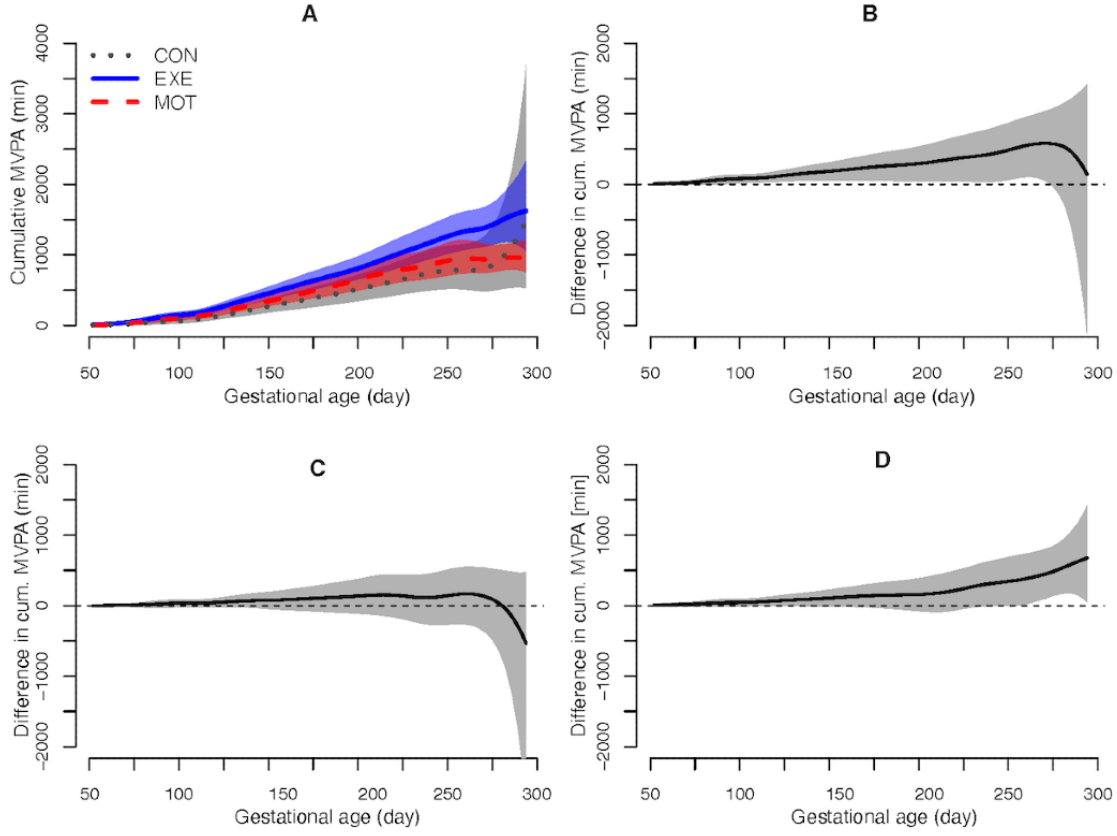
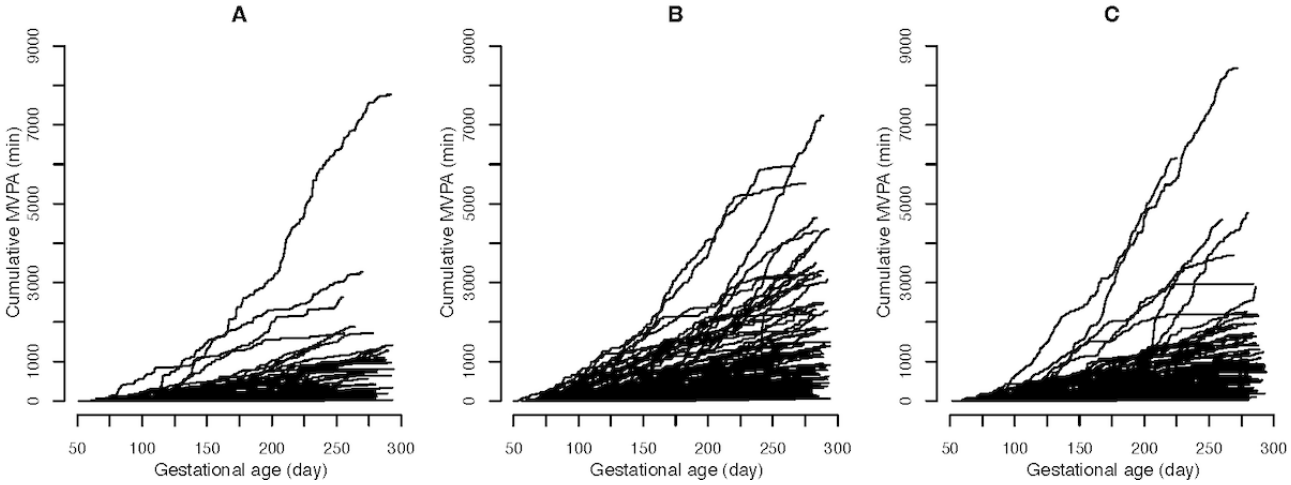


Figure 5. Individual cumulative moderate-to-vigorous intensity physical activity from randomization to delivery in (A) standard care, (B) structured supervised exercise training, and (C) motivational counseling on physical activity. MVPA: moderate-to-vigorous intensity physical activity.



COVID-19 Sensitivity Analysis

MVPA (minutes per week) did not differ between participants included before the COVID-19 pandemic (physical intervention only, 120/219) and during the COVID-19 pandemic (online intervention only, 63/219) in either CON (–14, 95% CI –49 to 22; *P*=.44), EXE (–16, 95% CI –42 to 11; *P*=.25), or MOT (–6, 95% CI –37 to 25; *P*=.712; Multimedia Appendix 2).

Women in EXE offered the online intervention only participated in more exercise sessions per week than women offered the physical intervention only (online: 1.6, 95% CI 1.3 to 2.0 and physical: 1.1, 95% CI 0.9 to 1.4; *P*=.03). Participants in EXE attended on average 4.9 swimming pool sessions during the online intervention period. The number of MOT sessions attended did not differ between women who were offered the intervention before or during the COVID-19 pandemic (physical: 5.3, 95% CI 4.6 to 6.0 and online: 5.6, 95% CI 4.8 to 6.4; *P*=.97). Participants included before the COVID-19 pandemic and delivered during (36/219) were excluded in this analysis based on their mixed intervention.

Secondary Activity Tracker Outcomes

All tracker outcomes are presented in Figure 3 and accompanying statistics in Multimedia Appendix 1. PA at a vigorous intensity (minutes per week) was higher in EXE than in both CON and MOT (CON vs EXE: randomization to visit 2: 13, 95% CI 4 to 22; randomization to delivery: 13, 95% CI 4 to 22; MOT vs EXE: randomization to visit 2: 9, 95% CI 1 to 16, randomization to delivery: 9, 95% CI 1 to 17). In addition, the maximum heart rate was 2 (95% CI 0.3 to 3) beats per minute higher in EXE compared with CON from randomization to visit 2. No other tracker outcomes differed between groups.

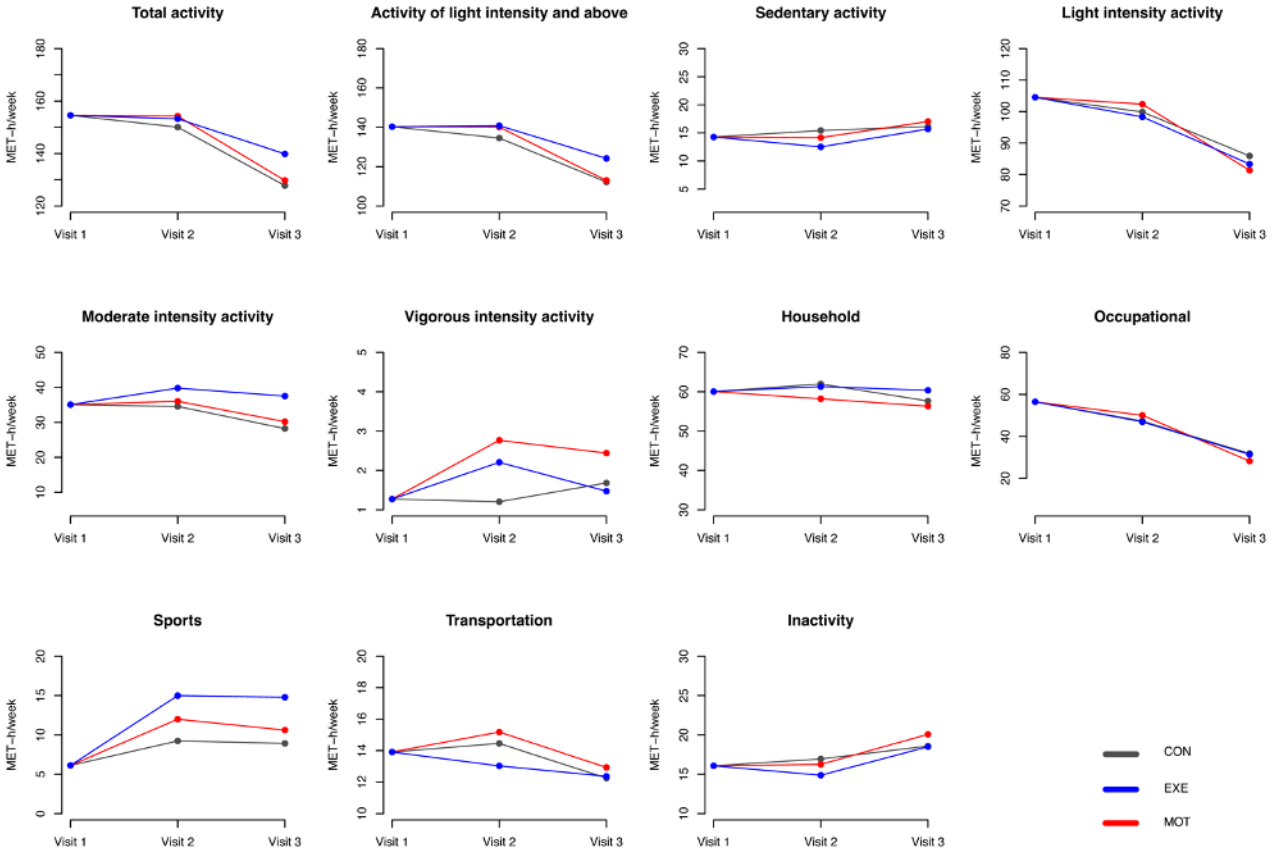
PA by PPAQ-DK

PPAQ-DK was completed for visits 1, 2, and 3 by 100% (219/219), 83.1% (182/219), and 77.2% (169/219) participants, respectively. Figure 6 shows the PA behaviors categorized by intensity and type. Differences between and within groups are shown in Multimedia Appendix 3 and Multimedia Appendix 4.

Total activity did not change from visit 1 to visit 2 in CON, EXE, or MOT, but PA decreased significantly from visit 1 to visit 3 in all groups (Multimedia Appendix 4). PA at moderate intensity was maintained at the same level over the course of pregnancy in CON, EXE, and MOT. However, participants in MOT increased PA at vigorous intensity from visit 1 to visit 2 and visit 1 to visit 3 (Multimedia Appendix 4). When combined (MVPA), the activity level (MET hours per week) did not change through pregnancy in any of the groups (CON: visit 1-2: –1, *P*=.90; visit 1-3: –4, *P*=.36; EXE: visit 1-2: 4, *P*=.10; visit 1-3: 1, *P*=.61; MOT: visit 1-2: 2, *P*=.40; visit 1-3: –5, *P*=.37; data not shown).

The MET hours per week spent at sports activities increased significantly from visit 1 to visit 2 and visit 1 to visit 3 for both EXE and MOT, while no changes were observed in CON (Multimedia Appendix 4). A comparison between groups revealed that sports was significantly higher in EXE compared with CON and MOT at both visit 2 and visit 3 (Multimedia Appendix 3).

Figure 6. Baseline-constrained comparison between groups based on the means of physical activity level from the Danish version of the Pregnancy Physical Activity Questionnaire. MET: metabolic equivalent of task; CON: standard care; EXE: structured supervised exercise training; MOT: motivational counseling on physical activity.



PA by DLW

A total of 134 participants (CON: 24/45, EXE: 53/87, MOT: 57/87) completed the DLW test and were included in the analysis. TEE (*P*=.14), AEE (*P*=.38), and PAL (*P*=.66) did not differ between groups (TEE [kcal per day]: CON 2215 [SD 238], EXE 2330 [SD 264], MOT 2331 [SD 260]; AEE [kcal per day]: CON 543 [SD 106], EXE 592 [SD 160], MOT 587 [SD 155]; and PAL [TEE/BMR]: CON 1.33 [SD 0.06], EXE 1.35 [SD 0.11], MOT 1.34 [SD 0.09]; Multimedia Appendix 5).

Adverse Events and Serious Adverse Events

Adverse events and serious adverse events from inclusion to delivery among all participants did not differ between groups (Multimedia Appendices 6-8).

Discussion

Principal Findings

FitMum aimed to investigate the effects of offering EXE or MOT to generate evidence about how to implement PA in healthy pregnant women's lives. We hypothesized that both EXE and MOT would increase MVPA in pregnancy compared with CON but that EXE would be more effective than MOT [33]. The study confirmed that EXE was more effective than CON, whereas MOT was not more effective than CON, and EXE and MOT did not differ. The number of adverse events and serious adverse events did not differ between groups.

Effectiveness of PA Interventions On PA Level In Pregnant Women

Several previous RCTs have used strategies like ours to examine how to increase PA in pregnant women and at the same time assessed the PA level by objective methods [13,24,26,44,45]. Seneviratne et al [24] conducted a 16-week stationary biking program in overweight and obese pregnant women and reported improved aerobic fitness compared to controls. When determining PA objectively by accelerometry, Hayman et al [26] found an immediate increase in MVPA after 4 weeks of tailored PA advice and access to a resource library. On the contrary, no increase in PA as determined by accelerometry was found after a combined aerobic and strength exercise program [44], face-to-face individual PA consultations [13], or app-based PA behavior change techniques [45].

Women in EXE were encouraged to participate in 3 hours of EXE per week, but the participants attended on average less than half of the sessions, and throughout their pregnancy, the MVPA level was only a third (54 of 150 minutes per week) of the internationally recommended amount [2]. As expected, MVPA was positively associated with the number of exercise sessions attended. Noticeably, EXE had a higher level of vigorous intensity PA compared with both CON and MOT. This was supported by a higher maximum heart rate among EXE.

Exercising at vigorous intensity is in accordance with recent suggestions for healthy pregnant women [46,47]. MOT had a high intervention attendance, but even though MOT contained face-to-face counseling, text messaging, activity tracker use, and behavior change techniques as recommended [13,48,49], we found no effect on MVPA compared with CON. The processes behind this finding are currently being assessed via mixed methods. The cumulative MVPA in EXE was significantly higher compared with MOT in the late part of pregnancy, and the same tendency was seen between CON and EXE. Interestingly, women who received the online EXE intervention due to COVID-19 restrictions joined 45% more exercise sessions compared with those who received the physical intervention.

Methodologies Used to Determine PA

Combining 3 different methodologies to assess PA using objective (activity tracker and DLW) and subjective (PPAQ-DK) methods provides insight into the complexity of PA. The activity tracker offers 24/7 measures of PA, and due to its convenience the tracker can be worn for a long period of time. However, commercial trackers are not designed for research purposes, and tracker algorithms are unknown. The PPAQ is considered one of the most valid and reliable questionnaires for the assessment of PA in pregnant women [27,50], but the inherent bias of self-reported PA is inevitable. The administration of the PPAQ-DK may have led to a heightened awareness of activity among participants [50], especially for members of the MOT group, who received a thorough review of their PA level at the counseling sessions. This might explain the perceived increase in vigorous intensity PA in MOT as determined by PPAQ-DK. DLW is the reference method for the determination of free-living energy expenditure and has previously been used to estimate PA level in pregnant women [39,51], but this is the first intervention study in pregnant women to include DLW. We found no significant differences between groups in TEE, AEE, or PAL, but this might be due to a lack of power, as TEE and AEE were 50 to 100 kcal per day higher in EXE and MOT compared with CON. On the other hand, active kilocalories recorded by the tracker and total activity obtained from the PPAQ-DK, which are equivalent to AEE from DLW, did not differ between groups. Therefore, the total activity probably did not differ between groups.

Strengths and Limitations

FitMum is the first RCT to compare the effectiveness of 2 different PA interventions in pregnant women. Strengths comprise the robust design based on the power of randomization, which leaves the internal validity high, and the comprehensive assessment of PA. The primary outcome was measured by a commercial activity tracker, which measured PA continuously, but no data on the validity of the tracker activity measurements has been published. The activity tracker may increase PA due to its motivational impact [49,52], but it might also not capture all activities. Notably, by default the tracker only reported activities with a MET value of ≥3 in bouts of at least 10 consecutive minutes as MVPA [38], and this might partly explain the relatively low MVPA in this study. An additional limitation was the impact of COVID-19 and the need to convert the physical interventions into online ones.

Conclusions

Findings from this RCT demonstrate that offering EXE is more effective than CON to implement MVPA in healthy pregnant women's lives. Offering MOT was not more effective than CON; EXE and MOT also did not differ. The MVPA in the intervention groups did not reach the recommended PA level in pregnancy. Changing the intervention to online due to COVID-19 restrictions did not affect MVPA level but increased exercise participation. Based on the most effective intervention on MVPA during pregnancy (EXE) and the increased level of EXE sessions attended in the online setup during the COVID-19 pandemic, it might be beneficial to add home-based, online exercise sessions in future prenatal PA interventions.

Acknowledgments

The authors would like to acknowledge all the women who participated in FitMum. Our thanks go to students, research assistants, and staff at the Department of Gynecology and Obstetrics, Copenhagen University Hospital–North Zealand, who took part in performing interventions and collecting data. Additionally, we would like to thank the technical staff at the Clinical Research Unit, Department of Clinical Research, Copenhagen University Hospital–North Zealand, especially Susanne Månsson and Charlotte Pietraszek, who engaged themselves in planning practicalities and collecting data. FitMum was funded by grant 8020-00353B from the Independent Research Fund Denmark, grant 128509 from TrygFonden, grant 061017 from the Copenhagen Center for Health Technology, grant 17-2-0883 from Beckett-Fonden, grant 10-002052 from the Aase and Ejnar Danielsens Fond, and grant 2017-1142 from the Familien Hede Nielsens Fond. In addition, funding was provided by the University of Copenhagen and Copenhagen University Hospital–North Zealand.

Authors' Contributions

BS initiated and directed FitMum. SdPK, CBR, JMB, TDC, SM, SAA, EL, and BS developed the study protocol. SdPK, CBR, ADJ, and IH conducted intervention activities and collected data assisted by SAA, research assistants, and master students. EL was the clinical trial manager and supervised the clinical part of FitMum in collaboration with JMB, TDC, SM, and BS. AKJ performed and supervised statistical analyses. SAA performed the activity tracker data management, and JEL contributed with expertise on self-tracking. GvH performed the doubly labeled water analysis. SdPK and SAA contributed equally, analyzed data, and wrote the manuscript. All authors read, contributed to, and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Comparison between groups based on imputed activity tracker datasets (intention-to-treat analysis) from randomization to visit 2 and delivery, respectively.
[DOCX File , 20 KB - jmir_v24i7e37699_app1.docx]

Multimedia Appendix 2
Moderate-to-vigorous intensity physical activity before and during the COVID-19 pandemic.
[PNG File , 59 KB - jmir_v24i7e37699_app2.png]

Multimedia Appendix 3
Pregnancy Physical Activity Questionnaire outcome differences.
[DOCX File , 19 KB - jmir_v24i7e37699_app3.docx]

Multimedia Appendix 4
Pregnancy Physical Activity Questionnaire outcome descriptive statistics.
[DOCX File , 19 KB - jmir_v24i7e37699_app4.docx]

Multimedia Appendix 5
One-way analysis of variance test of the doubly labeled water outcomes.
[PNG File , 128 KB - jmir_v24i7e37699_app5.png]

Multimedia Appendix 6
Summary of adverse events and serious adverse events.
[XLSX File (Microsoft Excel File), 9 KB - jmir_v24i7e37699_app6.xlsx]

Multimedia Appendix 7
All adverse events.
[XLSX File (Microsoft Excel File), 12 KB - jmir_v24i7e37699_app7.xlsx]

Multimedia Appendix 8
All serious adverse events.
[XLSX File (Microsoft Excel File), 10 KB - jmir_v24i7e37699_app8.xlsx]

References

1. ACOG. Physical activity and exercise during pregnancy and the postpartum period: ACOG Committee Opinion, Number 804. Obstet Gynecol 2020 Apr;135(4):e178-e188. [doi: 10.1097/AOG.0000000000003772] [Medline: 32217980]
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. Hayes L, McParlin C, Azevedo LB, Jones D, Newham J, Olajide J, et al. The effectiveness of smoking cessation, alcohol reduction, diet and physical activity interventions in improving maternal and infant health outcomes: a systematic review of meta-analyses. Nutrients 2021 Mar 23;13(3):1036 [FREE Full text] [doi: 10.3390/nu13031036] [Medline: 33806997]
4. Sánchez-Polán M, Franco E, Silva-José C, Gil-Ares J, Pérez-Tejero J, Barakat R, et al. Exercise during pregnancy and prenatal depression: a systematic review and meta-analysis. Front Physiol 2021 Jun 28;12:640024 [FREE Full text] [doi: 10.3389/fphys.2021.640024] [Medline: 34262468]
5. Kubler JM, Clifton VL, Moholdt T, Beetham KS. The effects of exercise during pregnancy on placental composition: a systematic review and meta-analysis. Placenta 2022 Jan;117:39-46. [doi: 10.1016/j.placenta.2021.10.008] [Medline: 34768167]
6. Kusuyama J, Alves-Wagner AB, Makarewicz NS, Goodyear LJ. Effects of maternal and paternal exercise on offspring metabolism. Nat Metab 2020 Sep 14;2(9):858-872 [FREE Full text] [doi: 10.1038/s42255-020-00274-7] [Medline: 32929233]
7. Bhattacharjee J, Mohammad S, Adamo KB. Does exercise during pregnancy impact organs or structures of the maternal-fetal interface? Tissue Cell 2021 Oct;72:101543. [doi: 10.1016/j.tice.2021.101543] [Medline: 33940567]

8. Bennett CJ, Walker RE, Blumfield ML, Ma J, Wang F, Wan Y, et al. Attenuation of maternal weight gain impacts infant birthweight: systematic review and meta-analysis. J Dev Orig Health Dis 2018 Nov 09;10(4):387-405. [doi: 10.1017/s2040174418000879]
9. Mottola MF, Davenport MH, Ruchat S, Davies GA, Poitras VJ, Gray CE, et al. 2019 Canadian guideline for physical activity throughout pregnancy. Br J Sports Med 2018 Nov 18;52(21):1339-1346. [doi: 10.1136/bjsports-2018-100056] [Medline: 30337460]
10. Díaz-Burrueco J, Cano-Ibáñez N, Martín-Peláez S, Khan KS, Amezcua-Prieto C. Effects on the maternal-fetal health outcomes of various physical activity types in healthy pregnant women: a systematic review and meta-analysis. Eur J Obstet Gynecol Reprod Biol 2021 Jul;262:203-215. [doi: 10.1016/j.ejogrb.2021.05.030] [Medline: 34058612]
11. Lindqvist M, Persson M, Mogren I. "Longing for individual recognition": pregnant women's experiences of midwives' counselling on physical activity during pregnancy. Sex Reprod Healthc 2018 Mar;15:46-53. [doi: 10.1016/j.srhc.2017.12.003] [Medline: 29389501]
12. Barakat R. An exercise program throughout pregnancy: Barakat model. Birth Defects Research 2020 Jul 02;113(3):218-226. [doi: 10.1002/bdr2.1747]
13. Currie S, Sinclair M, Murphy MH, Madden E, Dunwoody L, Liddle D. Reducing the decline in physical activity during pregnancy: a systematic review of behaviour change interventions. PLoS One 2013 Jun 14;8(6):e66385 [FREE Full text] [doi: 10.1371/journal.pone.0066385] [Medline: 23799096]
14. Williams SL, French DP. What are the most effective intervention techniques for changing physical activity self-efficacy and physical activity behaviour—and are they the same? Health Educ Res 2011 Apr 14;26(2):308-322. [doi: 10.1093/her/cyr005] [Medline: 21321008]
15. James P, Morgant R, Merviel P, Saraux A, Giroux-Metges M, Guillodo Y, et al. How to promote physical activity during pregnancy: a systematic review. J Gynecol Obstet Hum Reprod 2020 Nov;49(9):101864. [doi: 10.1016/j.jogoh.2020.101864] [Medline: 32663651]
16. Wang C, Wei Y, Zhang X, Zhang Y, Xu Q, Sun Y, et al. A randomized clinical trial of exercise during pregnancy to prevent gestational diabetes mellitus and improve pregnancy outcome in overweight and obese pregnant women. Am J Obstet Gynecol 2017 Apr;216(4):340-351. [doi: 10.1016/j.ajog.2017.01.037] [Medline: 28161306]
17. Hui AL, Back L, Ludwig S, Gardiner P, Sevenhuysen G, Dean HJ, et al. Effects of lifestyle intervention on dietary intake, physical activity level, and gestational weight gain in pregnant women with different pre-pregnancy Body Mass Index in a randomized control trial. BMC Pregnancy Childbirth 2014 Sep 24;14(1):331 [FREE Full text] [doi: 10.1186/1471-2393-14-331] [Medline: 25248797]
18. Gaston A, Prapavessis H. Using a combined protection motivation theory and health action process approach intervention to promote exercise during pregnancy. J Behav Med 2014 Apr 23;37(2):173-184. [doi: 10.1007/s10865-012-9477-2] [Medline: 23180287]
19. Ruart S, Sinnapah S, Hue O, Janky E, Antoine-Jonville S. Prenatal counseling throughout pregnancy: effects on physical activity level, perceived barriers, and perinatal health outcomes: a quasi-experimental study. Int J Environ Res Public Health 2020 Nov 29;17(23):8887 [FREE Full text] [doi: 10.3390/ijerph17238887] [Medline: 33260471]
20. Hawkins M, Chasan-Taber L, Marcus B, Stanek E, Braun B, Ciccolo J, et al. Impact of an exercise intervention on physical activity during pregnancy: the Behaviors Affecting Baby and You Study. Am J Public Health 2014 Oct;104(10):e74-e81. [doi: 10.2105/ajph.2014.302072]
21. Renault KM, Nørgaard K, Nilas L, Carlsen EM, Cortes D, Pryds O, et al. The Treatment of Obese Pregnant Women (TOP) study: a randomized controlled trial of the effect of physical activity intervention assessed by pedometer with or without dietary intervention in obese pregnant women. Am J Obstet Gynecol 2014 Feb;210(2):134-139. [doi: 10.1016/j.ajog.2013.09.029] [Medline: 24060449]
22. Barakat R, Perales M, Cordero Y, Bacchi M, Mottola M. Influence of land or water exercise in pregnancy on outcomes: a cross-sectional study. Med Sci Sports Exerc 2017;49(7):403. [doi: 10.1249/mss.0000000000001234]
23. Poston L, Bell R, Croker H, Flynn AC, Godfrey KM, Goff L, et al. Effect of a behavioural intervention in obese pregnant women (the UPBEAT study): a multicentre, randomised controlled trial. Lancet Diabetes Endocrinol 2015 Oct;3(10):767-777. [doi: 10.1016/s2213-8587(15)00227-2]
24. Seneviratne S, Jiang Y, Derraik J, McCowan L, Parry G, Biggs J, et al. Effects of antenatal exercise in overweight and obese pregnant women on maternal and perinatal outcomes: a randomised controlled trial. BJOG 2016 Mar 06;123(4):588-597. [doi: 10.1111/1471-0528.13738] [Medline: 26542419]
25. Pearce EE, Evenson KR, Downs DS, Steckler A. Strategies to promote physical activity during pregnancy: a systematic review of intervention evidence. Am J Lifestyle Med 2013 Jan 01;7(1):38-50 [FREE Full text] [doi: 10.1177/1559827612446416] [Medline: 24363633]
26. Hayman M, Reaburn P, Browne M, Vandelanotte C, Alley S, Short CE. Feasibility, acceptability and efficacy of a web-based computer-tailored physical activity intervention for pregnant women: the Fit4Two randomised controlled trial. BMC Pregnancy Childbirth 2017 Mar 23;17(1):96 [FREE Full text] [doi: 10.1186/s12884-017-1277-9] [Medline: 28335767]

27. Chasan-Taber L, Schmidt M, Roberts D, Hosmer D, Markenson G, Freedson P. Development and validation of a Pregnancy Physical Activity Questionnaire. Med Sci Sports Exerc 2004 Oct;36(10):1750-1760. [doi: 10.1249/01.mss.0000142303.49306.0d] [Medline: 15595297]
28. Krøner F, Knudsen SDP, Roland CB, Alomairah SA, Molsted S. Validity and reliability of the Danish version of the pregnancy physical activity questionnaire to assess levels of physical activity during pregnancy. J Matern Fetal Neonatal Med 2020 Dec 08:1-7. [doi: 10.1080/14767058.2020.1856807] [Medline: 33292038]
29. Wong W, Roberts S, Racette S, Das S, Redman L, Rochon J, et al. The doubly labeled water method produces highly reproducible longitudinal results in nutrition studies. J Nutr 2014 May;144(5):777-783 [FREE Full text] [doi: 10.3945/jn.113.187823] [Medline: 24523488]
30. Santini C, Imakawa T, Moisés E. Erratum: Physical activity during pregnancy: recommendations and assessment tools. Rev Bras Ginecol Obstet 2017 Oct 16;39(10):584-584 [FREE Full text] [doi: 10.1055/s-0037-1607299] [Medline: 29036753]
31. Abeysekera MV, Morris JA, O'Sullivan AJ. Techniques to measure free-living energy expenditure during pregnancy: a guide for clinicians and researchers. Obstet Med 2014 Mar 27;7(2):60-65. [doi: 10.1177/1753495x14528324]
32. Roland CB, Knudsen SDP, Alomairah SA, Andersen AD, Bendix J, Clausen TD, et al. Structured supervised exercise training or motivational counselling during pregnancy on physical activity level and health of mother and offspring: FitMum study protocol. BMJ Open 2021 Mar 19;11(3):e043671 [FREE Full text] [doi: 10.1136/bmjopen-2020-043671] [Medline: 33741668]
33. Statistical analysis plan (SAP) for the FitMum randomized controlled trial. 2020 Aug 05. URL: https://clinicaltrials.gov/ProvidedDocs/30/NCT03679130/SAP_000.pdf [accessed 2022-06-03]
34. Rotgers F. Motivational interviewing: preparing people to change addictive behavior. J Stud Alcohol 1993 Jul;54(4):507-507. [doi: 10.15288/jsa.1993.54.507a]
35. Markland D, Ryan RM, Tobin VJ, Rollnick S. Motivational interviewing and self–determination theory. J Soc Clin Psychol 2005 Sep;24(6):811-831. [doi: 10.1521/jscp.2005.24.6.811]
36. Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. Psychol Health 2011 Nov;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]
37. Recommendations for pregnant women. The Danish Health Authorities. 2021. URL: https://www.sst.dk/da/viden/fysisk-aktivitet/anbefalinger-om-fysisk-aktivitet/gravide [accessed 2022-06-03]
38. Garmin International. URL: https://www.garmin.com [accessed 2022-06-03]
39. Speakman JR, Yamada Y, Sagayama H, Berman ES, Ainslie PN, Andersen LF, IAEA DLW database group. A standard calculation methodology for human doubly labeled water studies. Cell Rep Med 2021 Feb 16;2(2):100203 [FREE Full text] [doi: 10.1016/j.xcrm.2021.100203] [Medline: 33665639]
40. Hronek M, Zadak Z, Hrnciarikova D, Hyspler R, Ticha A. New equation for the prediction of resting energy expenditure during pregnancy. Nutrition 2009 Sep;25(9):947-953. [doi: 10.1016/j.nut.2009.02.011] [Medline: 19477620]
41. Buuren SV, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. J Stat Soft 2011;45(3):1-67. [doi: 10.18637/jss.v045.i03]
42. Wood S. Generalized Additive Models: An Introduction with R, 2nd Edition. London: Routledge; 2017.
43. Koenker R, Ng P. Inequality constrained quantile regression. Sankhya Indian J Stat 2005;67(2):40. [doi: 10.1017/cbo9780511754098]
44. Oostdam N, van Poppel M, Wouters M, Eekhoff E, Bekedam D, Kuchenbecker W, et al. No effect of the FitFor2 exercise programme on blood glucose, insulin sensitivity, and birthweight in pregnant women who were overweight and at risk for gestational diabetes: results of a randomised controlled trial. BJOG 2012 Aug;119(9):1098-1107. [doi: 10.1111/j.1471-0528.2012.03366.x] [Medline: 22616913]
45. Sandborg J, Söderström E, Henriksson P, Bendtsen M, Henström M, Leppänen M, et al. Effectiveness of a smartphone app to promote healthy weight gain, diet, and physical activity during pregnancy (HealthyMoms): randomized controlled trial. JMIR Mhealth Uhealth 2021 Mar 11;9(3):e26091 [FREE Full text] [doi: 10.2196/26091] [Medline: 33704075]
46. Meah VL, Davies GA, Davenport MH. Why can't I exercise during pregnancy? Time to revisit medical 'absolute' and 'relative' contraindications: systematic review of evidence of harm and a call to action. Br J Sports Med 2020 Dec 08;54(23):1395-1404. [doi: 10.1136/bjsports-2020-102042] [Medline: 32513676]
47. Brown WJ, Hayman M, Haakstad LAH, Lamerton T, Mena GP, Green A, et al. Australian guidelines for physical activity in pregnancy and postpartum. J Sci Med Sport 2022 Jun;25(6):511-519. [doi: 10.1016/j.jsams.2022.03.008] [Medline: 35418334]
48. Leonard KS, Evans MB, Oravecz Z, Smyth JM, Symons Downs D. Effect of technology-supported interventions on prenatal gestational weight gain, physical activity, and healthy eating behaviors: a systematic review and meta-analysis. J Technol Behav Sci 2020 Aug 25;6(1):25-41. [doi: 10.1007/s41347-020-00155-6]
49. 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]

50. Sattler M, Jaunig J, Watson E, van Poppel M, Mokkink L, Terwee C, et al. Physical activity questionnaires for pregnancy: a systematic review of measurement properties. Sports Med 2018 Oct;48(10):2317-2346 [FREE Full text] [doi: 10.1007/s40279-018-0961-x] [Medline: 30094797]
51. Löf M. Physical activity pattern and activity energy expenditure in healthy pregnant and non-pregnant Swedish women. Eur J Clin Nutr 2011 Dec 27;65(12):1295-1301. [doi: 10.1038/ejcn.2011.129] [Medline: 21792212]
52. Laranjo L, Ding D, Heleno B, Kocaballi B, Quiroz JC, Tong HL, et al. Do smartphone applications and activity trackers increase physical activity in adults? Systematic review, meta-analysis and metaregression. Br J Sports Med 2021 Apr 21;55(8):422-432. [doi: 10.1136/bjsports-2020-102892] [Medline: 33355160]

Abbreviations

AEE: active energy expenditure
BMR: basal metabolic rate
CON: standard care
DLW: doubly labeled water
EXE: structured supervised exercise training
MET: metabolic equivalent of task
MOT: motivational counseling
MVPA: moderate-to-vigorous intensity physical activity
PA: physical activity
PAL: physical activity level
PPAQ-DK: Danish Pregnancy Physical Activity Questionnaire
RCT: randomized controlled trial
TEE: total energy expenditure

Edited by A Mavragani; submitted 05.03.22; peer-reviewed by K McDonnell, J Pugmire; comments to author 28.04.22; revised version received 18.05.22; accepted 31.05.22; published 20.07.22.

Please cite as:
Knudsen SDP, Alomairah SA, Roland CB, Jessen AD, Hergel IM, Clausen TD, Larsen JE, van Hall G, Jensen AK, Molsted S, Bendix JM, Løkkegaard E, Stallknecht B
Effects of Structured Supervised Exercise Training or Motivational Counseling on Pregnant Women's Physical Activity Level: FitMum - Randomized Controlled Trial
J Med Internet Res 2022;24(7):e37699
URL: https://www.jmir.org/2022/7/e37699
doi:10.2196/37699
PMID:35857356

Letter to the Editor

Clarity on the Type of Review. Comment on "Value Cocreation in Health Care: Systematic Review"

Fnu Kajal[1], MD, MSc
University of Arizona, Tucson, AZ, United States

Corresponding Author:
Fnu Kajal, MD, MSc
University of Arizona
Global Center
615 North Park Avenue
Tucson, AZ, 85721
United States
Phone: 1 7233866266
Email: kkajal@email.arizona.edu

Related Articles:

Comment on: http://www.jmir.org/2022/3/e33061/

Comment in: http://www.jmir.org/2022/7/e39397/

(J Med Internet Res 2022;24(7):e38457) doi:10.2196/38457

KEYWORDS
value cocreation; health care; patient value; health care professional value; systematic review

I have read the systematic review titled, "Value Cocreation in Health Care: Systematic Review," by Peng et al [1]. The objective of the paper was to identify and review the literature as the area of value cocreation is new to health care. The topic is very relevant as there is a need to add value to health care that will ultimately help to reduce health inequities.

While this review summarizes the literature well, it does not qualify as a systematic review. Foremost is the lack of a clear question the review seeks to answer. A systematic review is usually conducted to answer a question; in this case, the authors seem to have conducted a scoping or narrative review systematically.

The authors themselves state that this area of research is new and the literature is fragmented. Thus, it would have been better to have conducted a scoping review rather than a systematic review [2]. Further, the search terms for this review do not seem to be adequate to capture all research on the subject. For example, the phrases used in the search strategy do not include "respectful care," which is often used in value cocreation in health care systems.

In addition, a high-quality systematic review follows PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines and checklists. This must be addressed in reference to PICO (population, intervention, comparators, and outcome). The systematic review lacks clarity on comparators and does not provide a list of all outcomes for which data were sought [3]. The MMAT (Mixed Method Appraisal Tool) does mention the quality of studies but lacks the anticipated risk of bias assessment in individual studies. Further, the authors have also not detailed any variability between the studies through heterogeneity, which might have impacted the interpretation of the results [4].

Most problematic, however, is the framework developed and presented in this review. The methodology of mapping the findings onto an existing theory is not a standard method. The authors need to justify why this method was adopted. The utility of this framework, therefore, is also not clear.

This area of research is clearly very relevant, and the authors have tried to put together the literature on this, but their systematic review needs more details at the granular level for a better understanding of the gaps and solutions to address areas of concern in the future.

Conflicts of Interest

None declared.

References

1. Peng Y, Wu T, Chen Z, Deng Z. Value cocreation in health care: Systematic review. J Med Internet Res 2022 Mar 25;24(3):e33061 [FREE Full text] [doi: 10.2196/33061] [Medline: 35333183]
2. Peters M, 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]
3. Page M, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, 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]
4. Higgins J, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002 Jun 15;21(11):1539-1558. [doi: 10.1002/sim.1186] [Medline: 12111919]

Abbreviations

MMAT: Mixed Method Appraisal Tool
PICO: population, intervention, comparators, and outcome
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by T Leung; submitted 13.04.22; this is a non–peer-reviewed article;accepted 24.05.22; published 11.07.22.

Please cite as:
Kajal F
Clarity on the Type of Review. Comment on "Value Cocreation in Health Care: Systematic Review"
J Med Internet Res 2022;24(7):e38457
URL: https://www.jmir.org/2022/7/e38457
doi:10.2196/38457
PMID:35816377

Letter to the Editor

Authors' Reply to: Clarity on the Type of Review. Comment on "Value Cocreation in Health Care: Systematic Review"

Yuxin Peng[1], BA; Tailai Wu[1], PhD; Zhuo Chen[2,3], PhD; Zhaohua Deng[4], PhD

[1]School of Medicine and Health Management, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China
[2]Department of Health Policy and Management, College of Public Health, University of Georgia, Athens, GA, United States
[3]School of Economics, Faculty of Humanities and Social Sciences, University of Nottingham Ningbo China, Ningbo, China
[4]School of Management, Huazhong University of Science and Technology, Wuhan, China

Corresponding Author:
Tailai Wu, PhD
School of Medicine and Health Management
Tongji Medical College
Huazhong University of Science and Technology
13 Hangkong Road
Qiaokou District
Wuhan, 430030
China
Phone: 86 13477072665
Email: lncle2012@gmail.com

Related Articles:

Comment on: https://www.jmir.org/2022/6/e38457

Comment on: http://www.jmir.org/2022/3/e33061/

(J Med Internet Res 2022;24(7):e39397) doi:10.2196/39397

KEYWORDS

value cocreation; health care; patient value; health care professional value; systematic review

We thank Kajal [1] and the editors of the *Journal of Medical Internet Research* for providing this opportunity to discuss our paper [2] with an academic audience directly after the publication of our work.

Overall, we think our systematic review is not perfect, but we endeavor to bring contributions and values to health care knowledge. We believe our audience can find not only the flaws but also the values of our paper. We, along with the reviewers and editors of the *Journal of Medical Internet Research*, have worked together to make this systematic review as valuable as possible during the publication process; we hope the readers will benefit from it in their future studies.

Our specific responses to Kajal [1] are as follows: First, we believe our study is a systematic review rather than a scoping review since our review not only identified available studies but also identified principal results and areas for future research [3]. The integrative framework provided in our review could serve as the basis for decision-making in value cocreation in health care. We understand that scoping reviews and systematic reviews overlap with each other, but our review matches the methods of a systematic review. Moreover, if the audience read our paper more carefully, they will find that "this area of research is new, and literature is fragmented" is not our only motivation; we also propose other motivations, including "for VCCH, the factors are not explored systematically, underlying mechanisms of its factors are vague, and consequences are not fully investigated" [2]. Finally, we may not have formally proposed a research question in our review, but we did have a specific research aim with the following implied question: What are the dimensions, antecedents, and consequences of value cocreation in health care, and how do they relate?

Second, we think our current search terms are adequate for our review goals. We have tried other search terms related to our research topic, but not many related or qualified articles were found.

Third, the risk of biases and heterogeneity were assessed using the MMAT (Mixed Methods Appraisal Tool). This tool not only appraises the quality of individual studies given the heterogeneity of the study designs but also accounts for many biases including confounding bias, nonresponse bias, and sampling bias [4]. Meanwhile, many previous systematic reviews or systematic review protocols have used the MMAT

to assess the risk of bias, such as Xu et al [5], Pearson et al [6], and Gledhill et al [7].

Forth, we admit that developing and presenting a theoretical framework is not a standard method, but it is our unique way of contributing to knowledge in health care. As you can see in our paper, the framework could (1) map and visualize studies systematically, (2) provide a novel theoretical perspective, (3) and imply many future research directions directly. Regarding these 3 benefits, we believe it is necessary to present this framework even though it is not a standard method.

We hope our response has alleviated the concerns raised by Kajal [1].

Conflicts of Interest

None declared.

References

1. Kajal F. Clarity on the Type of Review. Comment on "Value Cocreation in Health Care: Systematic Review". J Med Internet Res 2022 Jun 27;24(7):e38457 [FREE Full text] [doi: 10.2196/38457]
2. Peng Y, Wu T, Chen Z, Deng Z. Value cocreation in health care: Systematic review. J Med Internet Res 2022 Mar 25;24(3):e33061 [FREE Full text] [doi: 10.2196/33061] [Medline: 35333183]
3. 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]
4. 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. EFI 2018 Dec 18;34(4):285-291. [doi: 10.3233/efi-180221]
5. Xu N, Lv A, Li T, Li X, Huang M, Su Y. Experiences of healthcare providers during the coronavirus pandemic and its impact on them: protocol for a mixed-methods systematic review. BMJ Open 2021 Feb 26;11(2):e043686 [FREE Full text] [doi: 10.1136/bmjopen-2020-043686] [Medline: 33637546]
6. Pearson SA, Taylor S, Marsden A, Yorke J. Access to systemic anti-cancer therapies for women with secondary breast cancer-protocol for a mixed methods systematic review. Syst Rev 2021 Jul 23;10(1):209 [FREE Full text] [doi: 10.1186/s13643-021-01761-y] [Medline: 34294151]
7. Gledhill A, Forsdyke D, Murray E. Psychological interventions used to reduce sports injuries: a systematic review of real-world effectiveness. Br J Sports Med 2018 Aug;52(15):967-971. [doi: 10.1136/bjsports-2017-097694] [Medline: 29463497]

Abbreviations

MMAT: Mixed Method Appraisal Tool

Edited by T Leung; submitted 09.05.22; this is a non–peer-reviewed article;accepted 24.05.22; published 11.07.22.

Please cite as:
Peng Y, Wu T, Chen Z, Deng Z
Authors' Reply to: Clarity on the Type of Review. Comment on "Value Cocreation in Health Care: Systematic Review"
J Med Internet Res 2022;24(7):e39397
URL: https://www.jmir.org/2022/7/e39397
doi:10.2196/39397
PMID:35816384

Corrigenda and Addenda

Correction: Health Perceptions and Misconceptions Regarding COVID-19 in China: Online Survey Study

Jiawei Zhou[1], MPA; Bishwajit Ghose[2], PhD; Ruoxi Wang[1], PhD; Ruijun Wu[3], PhD; Zhifei Li[3], PhD; Rui Huang[2], PhD; Da Feng[2], PhD; Zhanchun Feng[1], PhD; Shangfeng Tang[1], PhD

[1]School of Medicine and Health Management, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[2]School of Pharmacy, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[3]China National Center for Biotechnology Development, Beijng, China

Corresponding Author:
Shangfeng Tang, PhD
School of Medicine and Health Management
Tongji Medical College
Huazhong University of Science and Technology
13 Hangkong Road
Wuhan
China
Phone: 86 13349895639
Email: sftang2018@hust.edu.cn

Related Article:

Correction of: https://www.jmir.org/2020/11/e21099

(J Med Internet Res 2022;24(7):e39493) doi:10.2196/39493

In "Health Perceptions and Misconceptions Regarding COVID-19 in China: Online Survey Study" (J Med Internet Res 2020;22(11):e21099), the authors made six updates.

1. The originally published article indicated the selected provinces in Eastern, Central, and Western China in a map in Figure 1. As this information has already been explained in the text, the figure was deemed unnecessary and was removed in the corrected version of the article. Accordingly, the numbering and in-text citations of the remaining figures were also updated in the corrected article.

2. The original Figure 1's in-text citation appeared in the article as follows:

> *As seen in Figure 1, Hubei, Hunan, and Shanxi provinces were selected in Central China. In Eastern China, Guangdong, Zhejiang, and Fujian provinces were selected.*

In the corrected version, it has been changed to the following:

> *Hubei, Hunan, and Shanxi provinces were selected in Central China. In Eastern China, Guangdong, Zhejiang, and Fujian provinces were selected.*

3. The originally published Figure 2 (captioned "Participants' level of confidence in the success of the fight against the COVID-19 epidemic") is now renumbered as Figure 1.

4. The original Figure 2's in-text citation appeared in the article as follows:

> *In total, 83.2% of them believed that the Centers for Disease Control could do better in controlling the risk of recurrence, 84.3% thought that hospitals could do better in controlling the risk of recurrence, and only 20.2% thought that COVID-19 outbreaks could happen again (Figure 2), which is consistent with the high level of awareness of COVID-19 prevention measures among the participants.*

In the corrected version, it has been changed to the following:

> *In total, 83.2% of them believed that the Centers for Disease Control could do better in controlling the risk of recurrence, 84.3% thought that hospitals could do better in controlling the risk of recurrence, and only 20.2% thought that COVID-19 outbreaks could happen again (Figure 1), which is consistent with the high level of awareness of COVID-19 prevention measures among the participants.*

5. The originally published Figure 3 (captioned "Vulnerable populations and their risk of being misled by incorrect information in videos on social media. OR: odds ratio") is now renumbered as Figure 2.

6. The original Figure 3's in-text citation appeared in the article as follows:

> *The results of the stratified analysis showed that the population aged >60 years (OR 1.52, 95% CI 1.10-2.11), those with a lower- or middle-income level*

(OR 1.36, 95% CI 1.00-1.83), those who were not working and not able to work (OR 1.83, 95% CI 1.04-3.21), those with a household income <100,000 RMB (<US $14,954; OR 1.34, 95% CI 1.08-1.67), and those with >2 suspected symptoms (OR 2.95, 95% CI 1.50-5.80) were more likely to be misled by videos on social media (Figure 3).

In the corrected version, it has been changed to the following:

The results of the stratified analysis showed that the population aged >60 years (OR 1.52, 95% CI 1.10-2.11), those with a lower- or middle-income level (OR 1.36, 95% CI 1.00-1.83), those who were not working and not able to work (OR 1.83, 95% CI 1.04-3.21), those with a household income <100,000 RMB (<US $14,954; OR 1.34, 95% CI 1.08-1.67), and those with >2 suspected symptoms (OR 2.95, 95% CI 1.50-5.80) were more likely to be misled by videos on social media (Figure 2).

The correction will appear in the online version of the paper on the JMIR Publications website on July 13, 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 11.05.22; this is a non–peer-reviewed article;accepted 12.05.22; published 13.07.22.

Please cite as:
Zhou J, Ghose B, Wang R, Wu R, Li Z, Huang R, Feng D, Feng Z, Tang S
Correction: Health Perceptions and Misconceptions Regarding COVID-19 in China: Online Survey Study
J Med Internet Res 2022;24(7):e39493
URL: https://www.jmir.org/2022/7/e39493
doi:10.2196/39493
PMID:35830718

Short Paper

Use of Random Domain Intercept Technology to Track COVID-19 Vaccination Rates in Real Time Across the United States: Survey Study

Rikki H Sargent[1], PhD; Shaelyn Laurie[1], MMASc; Leo F Weakland[2], MS; James V Lavery[3,4], MSc, PhD; Daniel A Salmon[5,6,7], MPH, PhD; Walter A Orenstein[3,8], MD; Robert F Breiman[2,3,8], MD

[1]RIWI Corp, Toronto, ON, Canada

[2]Center for Global Health Innovation, Atlanta, GA, United States

[3]Hubert Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, GA, United States

[4]Center for Ethics, Emory University, Atlanta, GA, United States

[5]Institute for Vaccine Safety, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[6]Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[7]Department of Health, Behavior and Society, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[8]School of Medicine, Emory University, Atlanta, GA, United States

Corresponding Author:
Rikki H Sargent, PhD
RIWI Corp
180 Bloor Street West, Suite 1000
Toronto, ON, M5S2V6
Canada
Phone: 1 888 505 7494
Email: rsargent@riwi.com

Abstract

Background: Accurate and timely COVID-19 vaccination coverage data are vital for informing targeted, effective messaging and outreach and identifying barriers to equitable health service access. However, gathering vaccination rate data is challenging, and efforts often result in information that is either limited in scope (eg, limited to administrative data) or delayed (impeding the ability to rapidly respond). The evaluation of innovative technologies and approaches that can assist in addressing these limitations globally are needed.

Objective: The objective of this survey study was to assess the validity of Random Domain Intercept Technology (RDIT; RIWI Corp) for tracking self-reported vaccination rates in real time at the US national and state levels. RDIT—a form of online intercept sampling—has the potential to address the limitations of current vaccination tracking systems by allowing for the measurement of additional data (eg, attitudinal data) and real-time, rapid data collection anywhere there is web access.

Methods: We used RDIT from June 30 to July 26, 2021, to reach a broad sample of US adult (aged ≥18 years) web users and asked questions related to COVID-19 vaccination. Self-reported vaccination status was used as the focus of this validation exercise. National- and state-level RDIT-based vaccination rates were compared to Centers for Disease Control and Prevention (CDC)–reported national and state vaccination rates. Johns Hopkins University's and Emory University's institutional review boards designated this project as public health practice to inform message development (not human subjects research).

Results: By using RDIT, 63,853 adult web users reported their vaccination status (6.2% of the entire 1,026,850 American web-using population that was exposed to the survey). At the national level, the RDIT-based estimate of adult COVID-19 vaccine coverage was slightly higher (44,524/63,853, 69.7%; 95% CI 69.4%-70.1%) than the CDC-reported estimate (67.9%) on July 15, 2021 (ie, midway through data collection; $t_{63,852}$=10.06; P<.001). The RDIT-based and CDC-reported state-level estimates were strongly and positively correlated (r=0.90; P<.001). RDIT-based estimates were within 5 percentage points of the CDC's estimates for 29 states.

Conclusions: This broad-reaching, real-time data stream may provide unique advantages for tracking the use of a range of vaccines and for the timely evaluation of vaccination interventions. Moreover, RDIT could be harnessed to rapidly assess

demographic, attitudinal, and behavioral constructs that are not available in administrative data, which could allow for deeper insights into the real-time predictors of vaccine uptake–enabling targeted and timely interventions.

(J Med Internet Res 2022;24(7):e37920) doi:10.2196/37920

KEYWORDS

COVID-19; vaccination rates; Random Domain Intercept Technology; health technology; vaccination; vaccine tracker; web-based survey; epidemiology; health data; digital tool; online intercept sampling; health service

Introduction

Accurate and timely COVID-19 vaccination coverage data are vital for informing targeted, effective messaging and outreach and identifying barriers to equitable health service access. The tracking of vaccination rates is needed in conjunction with effective COVID-19 case surveillance and death monitoring, as well as research on predictors of vaccination uptake and vaccine hesitancy, to optimize vaccine coverage in persons at disproportionate risk of COVID-19 and severe outcomes [1-5]. However, gathering such vaccination rate data is challenging, even in high-income countries such as the United States. Bill Gates recently expressed the importance of investment in global pandemic preparedness [6], which should involve assessing and implementing surveillance systems that can be utilized globally.

In the United States, the Centers for Disease Control and Prevention (CDC) rely on multiple external sources for their vaccination tracker [7]. Although considered the gold standard, this approach is labor intensive and limited to the collection of administrative data. In other words, certain demographic, attitudinal, and behavioral constructs are not available. Another approach involves national polling by using traditional recruitment methods (eg, panels and random digit dialing), in which respondents self-report their vaccination status (eg, Kaiser Family Foundation surveys and the Census Household Pulse) [8,9]. These polls have the benefit of collecting extended information, which could allow for deeper insights into the factors that predict vaccine uptake. Although immensely valuable for assessing knowledge, attitudes, and practices, these polls often do not allow for large-scale, rapid data collection, nor do they provide real-time granular results (with data available daily on a variety of population subcategories). Further, neither of these approaches are easily expandable to other countries with differing monitoring systems and access.

To identify targetable risk factors, it is critical to combine the benefits of traditional survey research with the advantages of rapid and continuous data collection by using a method that is scalable and can be incorporated into a global surveillance system. As such, we assessed the validity of Random Domain Intercept Technology (RDIT; RIWI Corp)—a form of online intercept sampling [10]—for tracking self-reported vaccination rates in real-time, with broad reach, and at the national and state levels across the United States.

Methods

Procedure

We used RDIT from June 30 to July 26, 2021, to reach a broad sample of US web users. RDIT can be administered anywhere in the world where there is web access. When web users click on a registered but commercially inactive web link or type in a web address for a site that is dormant, they have a random chance of that link being temporarily managed by the company that owns and administers RDIT—RIWI Corp. In this situation, instead of coming across a "this page does not exist" notification, a survey is delivered. Web users then decide whether to anonymously participate and may exit the survey at any time. No incentives are provided for participation.

Upon encountering the landing page, web users chose their preferred language (English or Spanish), and were told that the survey was anonymous and about COVID-19. Web users who chose to participate first reported their age and gender (those aged under 18 years were immediately exited from the landing page). They then reported their vaccination status (the focus of the validation exercise). Following this, those who were vaccinated were administered questions about the vaccine they received, and those who were unvaccinated were given a series of questions about COVID-19 vaccines (intentions, incentives, barriers, etc). All respondents were asked if they had heard recent accounts of COVID-19 vaccination side effects and were asked to provide information on where they receive their news and entertainment. Except for age, gender, and race and ethnicity, the remaining demographic questions were asked at the conclusion of the survey (education, urban or rural living location, political affiliation, and annual household income).

Ethical Considerations

This project was approved as public health practice to inform message development (not human subjects research) by the institutional review boards of Johns Hopkins University and Emory University.

Measures

With regard to COVID-19 vaccination status, respondents were asked "Have you personally received the COVID-19 vaccine?" They were given the following four response options: "Yes, a single-dose vaccine (J&J)"; "Yes, the first of two doses (Moderna or Pfizer)"; "Yes, both doses of a two-dose vaccine (Moderna or Pfizer)"; and "I have not received a vaccine." RDIT captures nonpersonally identifiable state location information from a respondent's IP address, which is instantaneously translated into a unique identifier. For this validation analysis, we compared self-reported adult vaccination rates from RDIT (having received at least 1 dose) to those that were provided in the publicly available CDC *COVID-19 Vaccinations in the United States, Jurisdiction* data set on July 15, 2021 (midway through RDIT data collection) [11].

Statistical Analysis

RDIT-based vaccination rate estimates (and associated logit transformed 95% CIs) were computed for the full national sample (N=63,853) and by state among those with available state information (n=57,986). A series of correlational analyses and 1-sample *t* tests (2-tailed) were used to compare RDIT-based estimates to CDC-reported rates.

Results

Respondent Description

By using RDIT, 1,026,850 US web users were exposed to the survey. Of those, 63,853 (6.2%) were aged ≥18 years and reported their vaccination status. Respondents were distributed throughout the sampling window from June 30 to July 26, 2021, with an average of 2365 (SD 1078; range 566-4266; skew=0.04) respondents per day. With regard to gender, 42.4% (27,060/63,853) of respondents were women, 51.6% (32,939/63,853) were men, and 6% (3854/63,853) indicated "other" gender. The median age was 39 (range 18 to ≥85) years. Respondents with available state information were regionally representative (ie, the RDIT-based state sample size strongly and positively correlated with the census-based state population size [12]; r=0.98, 95% CI 0.96-0.99; P<.001).

Of the 46,955 respondents who reported their race and ethnicity (46,955/63,853, 73.5% of those who reported their vaccination status), 23,505 (50.1%) identified as White; 5702 (12.1%) identified as African American or Black; 5094 (10.8%) identified as Hispanic or Latinx; 3046 (6.5%) identified as Asian; 1367 (2.9%) identified as Native American, American Indian, or Alaskan Native; 4954 (10.6%) identified as multiracial; and 3287 (7%) identified as a nonspecified racial and ethnic group. Of the 14,801 respondents who reported their annual household income (14,801/63,853, 23.2% of those who reported their vaccination status), 6700 (45.3%) reported an income of US $50,000 or less, 4914 (33.2%) reported an income of US $50,001 to US $125,000, and 3187 (21.5%) reported an income of US $125,001 or more. The race and ethnicity and economic distributions observed among the respondents are comparable to the most recent (2020) estimates derived from the US Census Current Population Survey [13,14].

National Comparison

The national CDC-estimated vaccine coverage on July 15, 2021, was 67.9%. The RDIT-based estimate among the full sample (N=63,853) from June 30 to July 26, 2021, was slightly higher (44,524/63,853, 69.7%; 95% CI 69.4%-70.1%; $t_{63,852}$=10.06; P<.001).

State Comparison

The RDIT-based and CDC-reported state-level estimates were strongly and positively correlated (r=0.90, 95% CI 0.83-0.94; P<.001). RDIT-based estimates were higher than the CDC-reported estimates by a mean of 3% (SD 4.5%, 95% CI 1.7%-4.2%; t_{50}=4.71; P<.001). RDIT-based estimates were higher than the CDC's estimates for 37 states and were within 2 and 5 percentage points for 12 and 29 states, respectively. When considering the absolute value of the estimate discrepancies, states with more RDIT respondents were associated with smaller discrepancies (r=−0.43, 95% CI −0.64 to −0.18; P=.001). We observed the largest discrepancy for Alaska (percent difference: 14.5%), which was likely due in part to the small sample size (n=85). The states with RDIT estimates within 1 percentage point of the relative CDC estimate were California; Connecticut; Washington, District of Columbia; Florida; Maryland; Maine; New York; Texas; and Utah (Table 1).

Table 1. State-level Random Domain Intercept Technology (RDIT)–based and Centers for Disease Control and Prevention (CDC)–reported adult vaccination rates (at least 1 dose received).

State	Total RDIT respondents, N	RDIT vaccinated respondents, n	RDIT-based vaccination rate, % (95% CI)	CDC-reported vaccination rate, %	Difference[a], %
AK	85	66	77.6 (67.5-85.3)	63.1	14.5
AL	742	425	57.3 (53.7-60.8)	51.2	6.1
AR	407	248	60.9 (56.1-65.6)	54.2	6.7
AZ	1422	927	65.2 (62.7-67.6)	63.1	2.1
CA	6483	4930	76 (75-77.1)	76.2	−0.2
CO	1026	761	74.2 (71.4-76.8)	70.7	3.5
CT	556	441	79.3 (75.7-82.5)	80.3	−1
DC	270	201	74.4 (68.9-79.3)	73.8	0.6
DE	177	136	76.8 (70-82.5)	71.1	5.7
FL	5275	3518	66.7 (65.4-68)	66.1	0.6
GA	2054	1288	62.7 (60.6-64.8)	55.4	7.3
HI	264	202	76.5 (71-81.2)	84.1	−7.6
IA	420	297	70.7 (66.2-74.9)	64.5	6.2
ID	239	155	64.9 (58.6-70.7)	53.6	11.3
IL	2084	1483	71.2 (69.2-73.1)	73	−1.8
IN	1035	632	61.1 (58.1-64)	57.2	3.9
KS	442	298	67.4 (62.9-71.6)	63	4.4
KY	621	373	60.1 (56.2-63.9)	62.3	−2.2
LA	803	453	56.4 (53-59.8)	50	6.4
MA	1481	1189	80.3 (78.2-82.2)	83.2	−2.9
MD	1186	909	76.6 (74.1-79)	75.9	0.7
ME	182	144	79.1 (72.6-84.4)	78.4	0.7
MI	1658	1110	66.9 (64.6-69.2)	63.2	3.7
MN	903	665	73.6 (70.7-76.4)	70.7	2.9
MO	918	571	62.2 (59-65.3)	57	5.2
MS	355	197	55.5 (50.3-60.6)	47.7	7.8
MT	112	72	64.3 (55-72.6)	59.1	5.2
NC	1927	1253	65 (62.9-67.1)	60.4	4.6
ND	76	48	63.2 (51.7-73.3)	56.1	7.1
NE	338	219	64.8 (59.5-69.7)	65.9	−1.1
NH	181	141	77.9 (71.3-83.4)	74.3	3.6
NJ	1936	1454	75.1 (73.1-77)	77.4	−2.3
NM	346	251	72.5 (67.6-77)	77.7	−5.2
NV	717	492	68.6 (65.1-71.9)	63.2	5.4
NY	4350	3225	74.1 (72.8-75.4)	73.6	0.5
OH	1622	1069	65.9 (63.6-68.2)	59.8	6.1
OK	630	354	56.2 (52.3-60)	58	−1.8
OR	718	541	75.3 (72.1-78.4)	70.7	4.6
PA	1964	1421	72.4 (70.3-74.3)	76.7	−4.3
RI	313	228	72.8 (67.6-77.5)	76.9	−4.1

State	Total RDIT respondents, N	RDIT vaccinated respondents, n	RDIT-based vaccination rate, % (95% CI)	CDC-reported vaccination rate, %	Difference[a], %
SC	725	470	64.8 (61.3-68.2)	55.4	9.4
SD	109	73	67 (57.6-75.2)	64.9	2.1
TN	1206	728	60.4 (57.6-63.1)	53.5	6.9
TX	4837	3051	63.1 (61.7-64.4)	62.5	0.6
UT	529	349	66 (61.8-69.9)	66.3	−0.3
VA	1855	1426	76.9 (74.9-78.7)	72.1	4.8
VT	67	61	91 (81.4-95.9)	85.9	5.1
WA	1292	946	73.2 (70.7-75.6)	75.5	−2.3
WI	731	523	71.5 (68.2-74.7)	66.2	5.3
WV	231	146	63.2 (56.8-69.2)	54.5	8.7
WY	86	50	58.1 (47.4-68.1)	50.9	7.2

[a]Difference = RDIT-based vaccination rate − CDC-reported vaccination rate.

Discussion

Principal Findings

RDIT provided similar estimates to those provided via the CDC method (the current standard) for vaccination rates at the national and state levels; however, estimates from RDIT are accessible daily at variable magnitudes and have region-targeting capabilities. Although the July 2021 RDIT-based national estimate was higher than the CDC estimate by 2%, it was more comparable to the CDC estimate than those derived from the Census Household Pulse and Delphi-Facebook surveys, which overestimated vaccination coverage by 17% and 14% in May 2021, respectively [15]. At the state level, RDIT-estimated rates strongly correlated with CDC reports, were slightly higher on average, and were within a 5% margin for 57% (29/51) of the states. These findings provide early evidence of the validity of RDIT as a complementary surveillance mechanism for tracking COVID-19 vaccination coverage across the United States.

Limitations

There are limitations to RDIT that should be considered. First, RDIT only reaches the web-using population; nonetheless, RDIT reaches a diverse set of the web-using population, including respondents who are not habitual survey takers, thereby allowing for subgroup analyses as needed. Second, a repeated measures assessment (ie, a follow-up to assess changes in vaccine status per individual) is not possible because RDIT does not collect identifying information. However, population-level changes can be identified. Similarly, without identifying information, opt-in bias is unknown, but one can record and evaluate trends in retention throughout the survey. Additionally, while associated with analytic limitations, the anonymous nature of RDIT is a strength, as participant privacy is prioritized. Although RDIT enables the collection of additional factors that may provide targets for improving vaccine coverage, it is possible for participants to drop out before they provide this information. An analysis of such drop-off can provide further understanding, and researchers can decide whether it is most appropriate to draw insights from all available data or only from respondents who complete the entire question set (in this case, we chose the former). Finally, administrative data, such as the data that inform the CDC vaccination tracker, could be less susceptible to self-reporting bias; however, we found strong correlations between the self-reported vaccination rate estimates and the CDC administrative metrics.

Conclusion

Access to this broad-reaching data stream is potentially less labor intensive than the alternative approaches that are currently used by the CDC, and RDIT-based estimates demonstrate adequate accuracy when compared to CDC estimates. RDIT's real-time nature may be a valuable tool for tracking vaccine uptake; pinpointing localities for targeted, timely interventions; and enabling the rapid evaluation of interventions and messaging campaigns globally. Of course, further investigation is needed to assess the accuracy of RDIT for tracking vaccination status globally. Nonetheless, RDIT could be harnessed to rapidly assess demographic, attitudinal, and behavioral constructs that are not available in administrative data, which could allow for deeper insights into the real-time predictors of COVID-19 vaccine uptake. Such data could be translated into effective interventions to strengthen vaccination coverage.

Acknowledgments

This project was made possible by a grant from the US Centers for Disease Control and Prevention, in collaboration with the Task Force for Global Health, under award number 6 NU51IP000873-05-02.

Conflicts of Interest

WAO is on the scientific advisory board for Moderna. JVL is a member of the Bioethics Advisory Council for Pfizer Inc. DAS is a member of the Janssen Policy Board and receives consulting and grant funding from Merck. SL is an employee of, and RHS is a consultant to, RIWI Corp—the company that owns the technology that was used to conduct the surveys. No other authors have conflicts of interest to report.

References

1. Hennessee I, Clennon JA, Waller LA, Kitron U, Bryan JM. Considerations for improving reporting and analysis of date-based COVID-19 surveillance data by public health agencies. Am J Public Health 2021 Dec;111(12):2127-2132. [doi: 10.2105/AJPH.2021.306520] [Medline: 34878867]
2. Ahmad FB, Anderson RN, Knight K, Rossen LM, Sutton PD. Advancements in the National Vital Statistics System to meet the real-time data needs of a pandemic. Am J Public Health 2021 Dec;111(12):2133-2140. [doi: 10.2105/AJPH.2021.306519] [Medline: 34878853]
3. Khairat S, Zou B, Adler-Milstein J. Factors and reasons associated with low COVID-19 vaccine uptake among highly hesitant communities in the US. Am J Infect Control 2022 Mar;50(3):262-267 [FREE Full text] [doi: 10.1016/j.ajic.2021.12.013] [Medline: 34995722]
4. Ndugga N, Hill L, Artiga S, Haldar S. Latest data on COVID-19 vaccinations by race/ethnicity. Kaiser Family Foundation. 2022 Apr 07. URL: https://www.kff.org/coronavirus-covid-19/issue-brief/latest-data-on-covid-19-vaccinations-by-race-ethnicity/ [accessed 2022-05-31]
5. Sargent RH, Laurie S, Moncada L, Weakland LF, Lavery JV, Salmon DA, et al. Masks, money, and mandates: A national survey on efforts to increase COVID-19 vaccination intentions in the United States. PLoS One 2022 Apr 21;17(4):e0267154. [doi: 10.1371/journal.pone.0267154] [Medline: 35446922]
6. Neville S, Kuchler H. Bill Gates calls for global surveillance team to spot pandemic threats. Financial Times. 2022 May 01. URL: https://www.ft.com/content/c8896c10-35da-46aa-957f-cf2b4e18cfce [accessed 2022-05-31]
7. COVID-19 vaccination data in the United States. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/vaccines/distributing/about-vaccine-data.html [accessed 2021-10-27]
8. KFF COVID-19 vaccine monitor archives. Kaiser Family Foundation. URL: https://www.kff.org/coronavirus-covid-19/kff-covid-19-vaccine-monitor-archives/ [accessed 2021-11-08]
9. Measuring household experiences during the coronavirus pandemic. United States Census Bureau. URL: https://www.census.gov/data/experimental-data-products/household-pulse-survey.html [accessed 2021-12-20]
10. Schonlau M, Couper MP. Options for conducting web surveys. Stat Sci 2017 May;32(2):279-292 [FREE Full text] [doi: 10.1214/16-sts597]
11. COVID-19 vaccinations in the United States, jurisdiction. Centers for Disease Control and Prevention. URL: https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-Jurisdi/unsk-b7fc [accessed 2022-01-04]
12. State population totals and components of change: 2020-2021. United States Census Bureau. URL: https://www.census.gov/data/tables/time-series/demo/popest/2020s-state-total.html [accessed 2021-12-22]
13. Population distribution by race/ethnicity (CPS). Kaiser Family Foundation. URL: https://tinyurl.com/3h874zkp [accessed 2022-05-31]
14. HINC-01. Selected characteristics of households by total money income. United States Census Bureau. URL: https://www.census.gov/data/tables/time-series/demo/income-poverty/cps-hinc/hinc-01.html [accessed 2022-05-31]
15. Bradley VC, Kuriwaki S, Isakov M, Sejdinovic D, Meng XL, Flaxman S. Unrepresentative big surveys significantly overestimated US vaccine uptake. Nature 2021 Dec;600(7890):695-700 [FREE Full text] [doi: 10.1038/s41586-021-04198-4] [Medline: 34880504]

Abbreviations

CDC: Centers for Disease Control and Prevention
RDIT: Random Domain Intercept Technology

Edited by A Mavragani; submitted 11.03.22; peer-reviewed by R Schechter, N Seeman; comments to author 08.05.22; revised version received 06.06.22; accepted 07.06.22; published 01.07.22.

Please cite as:
Sargent RH, Laurie S, Weakland LF, Lavery JV, Salmon DA, Orenstein WA, Breiman RF
Use of Random Domain Intercept Technology to Track COVID-19 Vaccination Rates in Real Time Across the United States: Survey Study
J Med Internet Res 2022;24(7):e37920
URL: https://www.jmir.org/2022/7/e37920
doi:10.2196/37920
PMID:35709335

Original Paper

Studying the Effect of Long COVID-19 Infection on Sleep Quality Using Wearable Health Devices: Observational Study

Mario Mekhael[1], MD, MSc; Chan Ho Lim[1], BSc; Abdel Hadi El Hajjar[1], MD; Charbel Noujaim[1], MD, MSc; Christopher Pottle[1], MSc; Noor Makan[1], BSc; Lilas Dagher[2], MD; Yichi Zhang[1], BSc; Nour Chouman[1], MD; Dan L Li[1], MD; Tarek Ayoub[1], MD; Nassir Marrouche[1], MD

[1]Tulane University School of Medicine, New Orleans, LA, United States

[2]Department of Medicine, Emory University School of Medicine, Atlanta, GA, United States

Corresponding Author:
Nassir Marrouche, MD
Tulane University School of Medicine
Suite A128
1324 Tulane Ave
New Orleans, LA, 70112
United States
Phone: 1 5049883072
Email: nmarrouche@tulane.edu

Abstract

Background: Patients with COVID-19 have increased sleep disturbances and decreased sleep quality during and after the infection. The current published literature focuses mainly on qualitative analyses based on surveys and subjective measurements rather than quantitative data.

Objective: In this paper, we assessed the long-term effects of COVID-19 through sleep patterns from continuous signals collected via wearable wristbands.

Methods: Patients with a history of COVID-19 were compared to a control arm of individuals who never had COVID-19. Baseline demographics were collected for each subject. Linear correlations among the mean duration of each sleep phase and the mean daily biometrics were performed. The average duration for each subject's total sleep time and sleep phases per night was calculated and compared between the 2 groups.

Results: This study includes 122 patients with COVID-19 and 588 controls (N=710). Total sleep time was positively correlated with respiratory rate (RR) and oxygen saturation (SpO_2). Increased awake sleep phase was correlated with increased heart rate, decreased RR, heart rate variability (HRV), and SpO_2. Increased light sleep time was correlated with increased RR and SpO_2 in the group with COVID-19. Deep sleep duration was correlated with decreased heart rate as well as increased RR and SpO_2. When comparing different sleep phases, patients with long COVID-19 had decreased light sleep (244, SD 67 vs 258, SD 67; *P*=.003) and decreased deep sleep time (123, SD 66 vs 128, SD 58; *P*=.02).

Conclusions: Regardless of the demographic background and symptom levels, patients with a history of COVID-19 infection demonstrated altered sleep architecture when compared to matched controls. The sleep of patients with COVID-19 was characterized by decreased total sleep and deep sleep.

(J Med Internet Res 2022;24(7):e38000) doi:10.2196/38000

KEYWORDS

COVID-19; digital health; wearables; sleep; long COVID-19; wearable device; demographic; biometric; patient data; sleep architecture; health data; health monitoring

Introduction

Although COVID-19 is primarily known as a pulmonary disease [1], literature suggests significant consequences regarding daily activities and mental health due to the infection itself or associated quarantine [2]. Moreover, reports indicate increased incidences of psychologic and psychiatric conditions during the pandemic such as sleep disturbances and decreased accessibility

to health care [3,4]. Thus, there is a need for remote continuous monitoring, telemedicine, and digital health monitoring systems to bridge the gap between patients and physicians [5].

The sleep cycle is traditionally divided in two phases: rapid eye movement (REM) sleep and nonrapid eye movement (NREM). Furthermore, NREM sleep is divided into the three subphases of (1) awake, (2) light, and (3) deep sleep. Human body usually cycles through these phases 4 to 6 times per night with 90 minutes in each stage [6]. Initially, sleep has been studied using polysomnography, which is a multisensor system that has been the gold standard for analyzing sleep stages and sleep-related disorders [7]. However, polysomnography has many drawbacks, such as the need for a hospital stay and its high-cost logistics such as the use of complex hardware needed for electroencephalographic, electromyographic, and electrooculographic assessments. All those factors can alter physiological sleep architecture and bias the results. Consequently, less than half of sleep studies nowadays are conducted in formal sleep facilities [8]. Having said that, wearable technology has been developed in the last decade, which consists of smart devices or gadgets worn close to or in contact with the skin used to capture biometric data [9]. With the recent trend of wearables, we have seen the development of photoplethysmography (PPG) technology to analyze different sleep phases, avoiding challenges that accompany the traditional polysomnography exam. In fact, reflective light emitted by the wearable device allows to measure blood volume changes in the vessels, which allows for the accurate measurement of heart rate (HR) and heart rate variability (HRV) [10]. HRV serves as a surrogate to estimate the effect of both sympathetic and parasympathetic nervous systems on the cardiovascular system. In addition, activities of both parasympathetic and sympathetic nervous systems vary in different sleep phases. For example, increased parasympathetic nervous system activity and therefore decreased HR was noticed in deeper stages of sleep [11,12]. Consequently, machine learning algorithms have been developed using the relationship between biometrics (such as HR and HRV) and sleep cycle to define sleep phases using PPG [13-16].

Long COVID-19 syndrome is defined as symptoms that persist after acute COVID-19 infection; however, the definitions vary by literature [17-19]. Previous studies have shown increased sleep disturbances and decreased sleep quality during and after COVID-19 infection [20,21]. However, those studies focused on qualitative analyses based on subjective measurements and survey responses rather than quantitative data [22,23]. Hence, in this paper, we study and evaluate the long-term effects of COVID-19 on sleep patterns using the continuously monitored metrics from wristband devices.

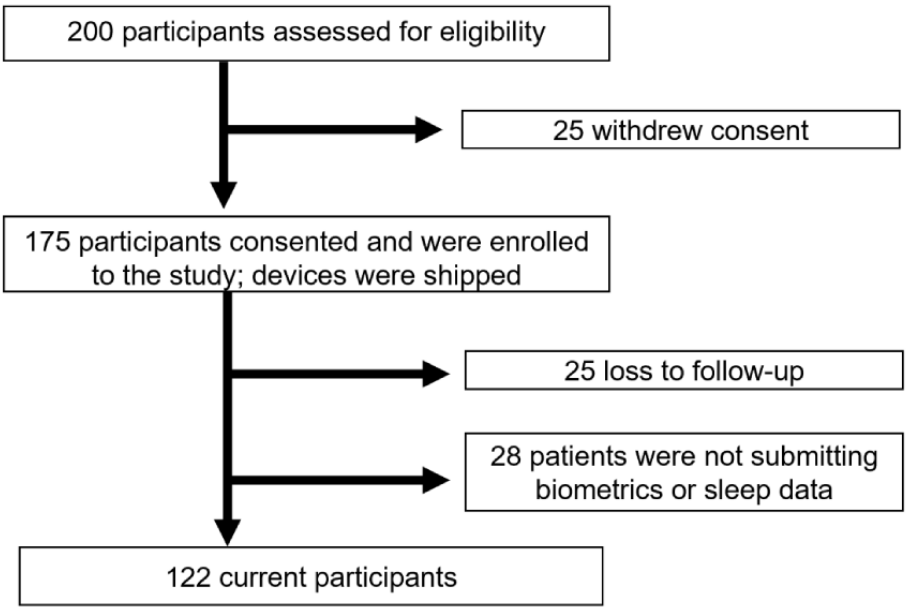
Methods

Study Design

Wearables to Investigate the Long Term Cardiovascular and Behavioral Impacts of COVID-19 (WEAICOR) is a prospective observational study of subjects 18 years or older who were monitored using the Biostrap wearable or wristband device. The study aims to identify the impact of long COVID-19 infection on sleep using wearables. In this analysis, we sought to compare continuous data recorded using a wearable device between patients who were diagnosed and recovered from COVID-19 and controls who were never diagnosed with the disease.

Patient's enrollment flowchart is represented in Figure 1. After eligibility screening and signing electronic consent forms, all subjects were sent a Biostrap device by mail to continuously monitor their biometric data. Biometric parameters included HR, HRV, respiratory rate (RR), and oxygen saturation (SpO_2). Device Instructions tailored to the study were provided by phone call by the study coordinator, along with a recorded video detailing the steps to activate the device with the mobile app.

Figure 1. Study flowchart.



Ethics Approval

WEAICOR study was approved by Tulane University Institution Review Board on June 09, 2020 (Study #2020-678).

Study Population

In this analysis, patients who had COVID-19 and recovered (study arm) will be compared to a control arm of participants who never had COVID-19 or associated symptoms. The study arm recruitment was carried out through flyers and advertisements on different platforms of social media along with mass emails generated to the Tulane staff and student body. A total of 200 participants were assessed for eligibility by September 2021. The control data were collected from a group of participants who opted in to an internal Biostrap study from April 12, 202, to July 31, 2020, as a part of their COVID-19 initiative. The participants received a baseline questionnaire collecting demographic and medical history data. Additionally, a daily survey was sent to all individuals to identify any COVID-19 symptom or positive COVID-19 case in each participant's household. Only individuals who consistently answered "No" regarding a positive COVID-19 diagnosis and denied related symptoms were included in the control group. Additionally, the existing users were willing to contribute their deidentified data for research. We secured informed consent forms and listed Tulane University as an organization with data access.

Biostrap Device

Biostrap is a PPG-based smartband that records patients' vitals at rest with 5-minute intervals and generates graphic results and reports on the Biostrap mobile app. PPG is an optical technique for detecting blood volume changes within the blood vessels by the changes in the light received from the photodiode to estimate physiological parameters. Biometrics such as HR, HRV, RR, and SpO_2 along with others related to the cardiovascular and autonomic nervous systems can be computed noninvasively using collected infrared signals. When paired with infrared signals, a red-light signal enables SpO_2 estimation. The combination of all those parameters along with arm movement enables us to classify sleep cycle into the 3 different phases of awake sleep, light sleep, and deep sleep. Example of biometric recordings (Figure 2) and sleep analysis recordings (Figure 3) are provided for simplification. Figure 2 shows biometrics recordings during a single night for a patient with COVID-19. Figure 3 describes the summary report and time spent in different sleep phases in a single night for a patient with COVID-19. PPG and accelerometer data collected from the wrist are transferred by the mobile app to the Biostrap cloud server, where they undergo signal processing and machine learning algorithms to generate physiological data at rest and transfer it to Tulane's data warehouse server. The accuracy and reproducibility of the Biostrap device in assessing basic physiological data have already been reported in previously published studies [24,25].

Figure 2. Recording example of biometrics during a night for a patient with long COVID-19. (a) RESP: respiratory rate (respirations per minute); (b) SpO_2: saturation of oxygen (%); (c) HR: heart rate (beats per minute); and (d) HRV: heart rate variability (beats per minute).

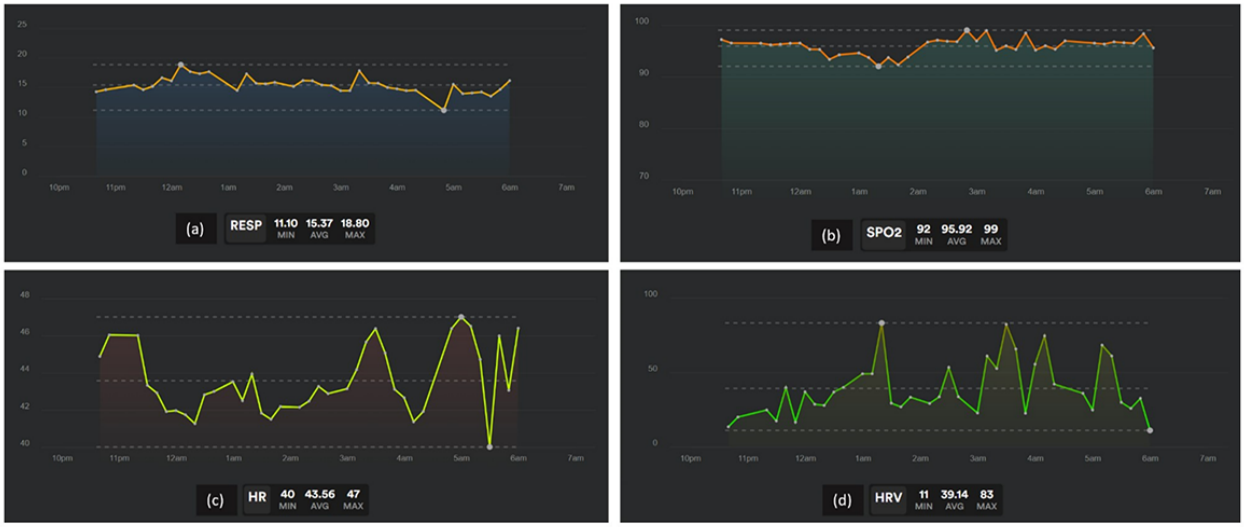
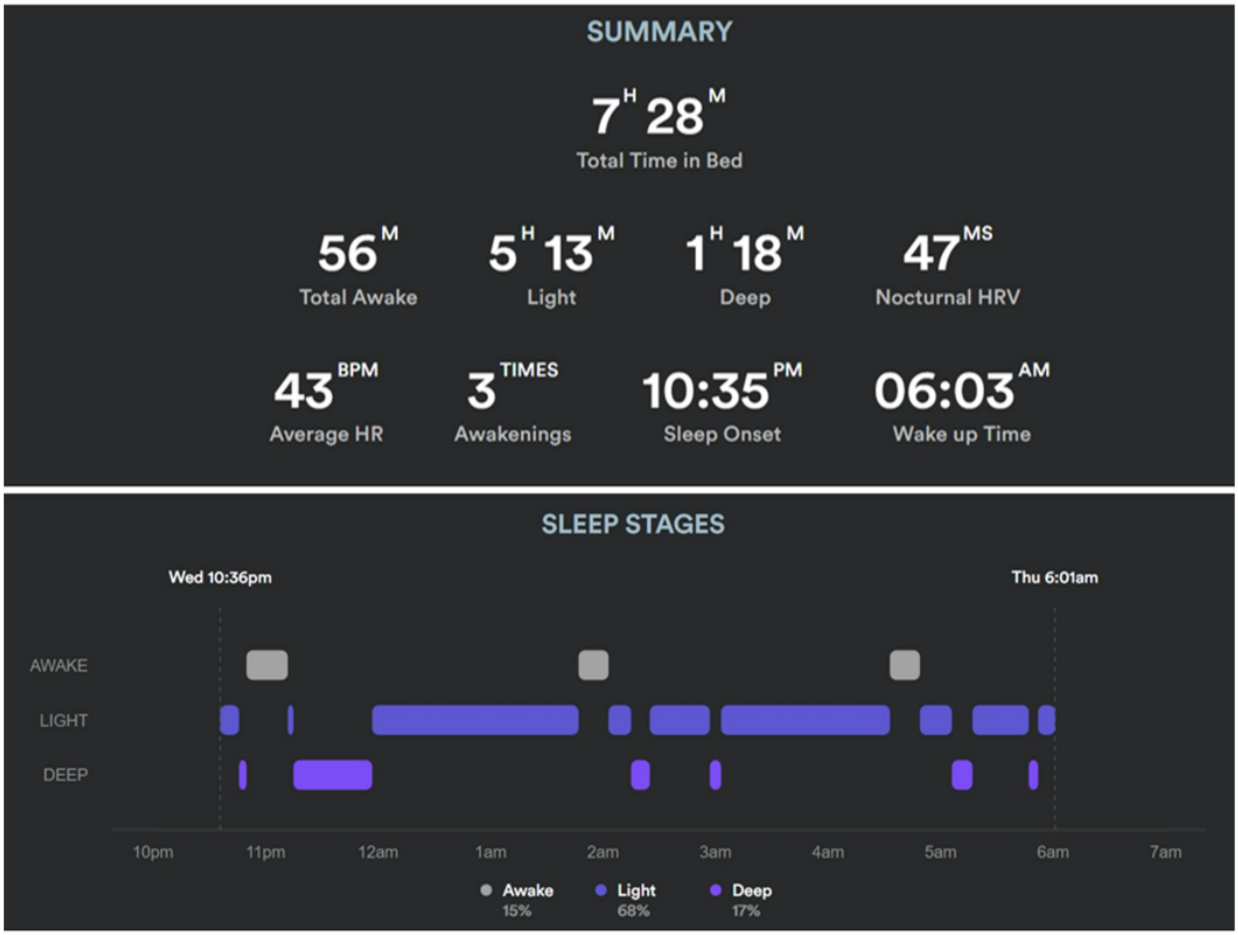


Figure 3. Recording example of sleep summary and sleep phases during a night for a patient with long COVID-19. HR: heart rate; HRV: heart rate variability.



Patient Follow-up

The research nurse and coordinator assigned to the study used Biostrap's remote data to ensure adequate data generation, patient compliance, and technical troubleshooting. Throughout the study, they actively followed up with the patients through phone calls and emails to address any problem or concern.

Data Analysis

The following baseline characteristics were collected for each participant: age, gender, BMI, comorbidities, educational level, and COVID-19 symptoms severity level. The means of these baseline characteristics were calculated for COVID-19 and controls and compared using 2-sample *t* test (2-tailed).

For each participant, the average duration in different sleep phases per night (awake, light, deep, and total) was calculated. For each sleep phase, we took the number of minutes per phase per day and calculated the average over the total number of days. Pearson bivariate linear correlations among the mean duration in sleep phases and means of biometrics (HR, HRV, RR, and SpO_2) were performed for COVID-19, controls, and the whole study population (specified as "Cohort" in the results section) to evaluate the association between the different components of the autonomic system and sleep cycle (Figure 4). The direction of changes in biometrics and duration of sleep phases will allow us to understand more the interaction between these 2 systems.

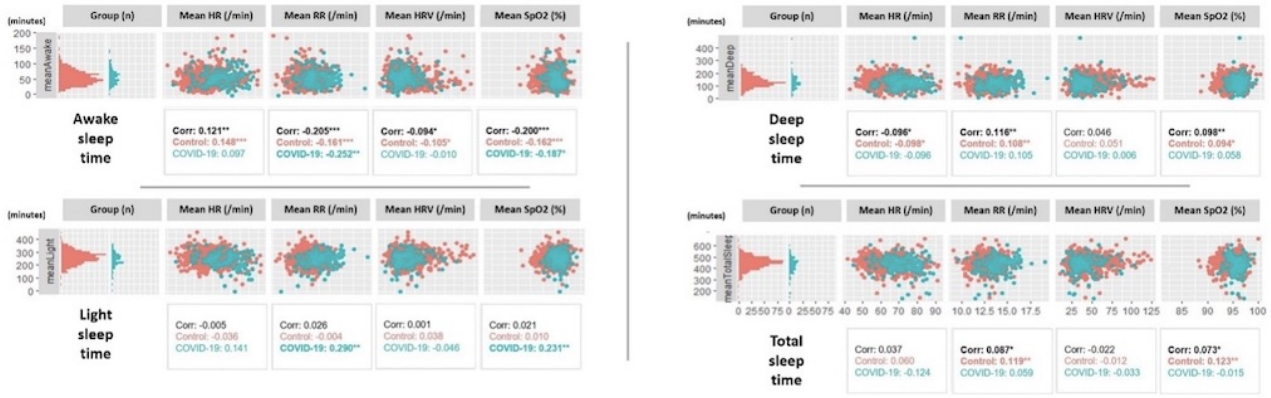
In addition, the participants from the whole cohort were divided into groups depending on their biometrics average during sleep (higher HR: >80 beats per minute vs lower HR: <80 beats per minute [26]; higher HRV: >20 milliseconds vs lower HRV: <20 milliseconds; and higher RR: >20 breaths per minute vs lower RR: <20 breaths per minute). Sleep phases (total, awake, light, and deep sleep) between the different groups were compared using Mann-Whitney *U* test. Patients with low HRV (n=27) and high RR (n=0) were little in number, and therefore the analysis was not statically significant for HRV and was not feasible for RR.

The mean measurements for each participant's total sleep time and sleep phases per night (awake, light, deep, and total) were calculated and weighted proportionally to the number of days each participant submitted data. For example, a participant with 25 nights of sleep data would have half of the weight of one with 50 nights of sleep data. From that weighted set, the median of each group's sleep times was taken and recorded, as the distributions of mean sleep times across both groups departed significantly from normality according to the Shapiro Wilk test. Distributions of sleep times in the control group and the group with COVID-19 were compared using the 2-Sample Wilcoxon (Mann-Whitney *U*) test. The same analysis was conducted in

an unweighted manner, where, for example, a subject with 25 nights of sleep data had just as much an effect on the test as a subject with 50 nights of sleep data. Two-sided *P* values of less than .05 were considered significant.

To mitigate potential selection bias arising from the observational nonrandomized study design, we applied propensity score matching and achieved a more balanced control group. During the matching process, the study participants with a history of COVID-19 were matched 1:1 to the participants without any history of COVID-19 by calculating the propensity score of the participants having COVID-19. The propensity score was carried out by estimating the probability of each participant having a history of COVID-19 based on age, BMI, and gender through multivariable logistic classifier.

Figure 4. Correlations between different sleep phases and biometrics. Corr: correlation with the whole cohort; HR: heart rate; HRV: heart rate variability; RR: respiratory rate; SpO_2: oxygen saturation. **P*<.05; ***P*<.01; ****P*<.001.



Results

Baseline Characteristics

We included 122 patients in the group with COVID-19 and 588 participants in the control group. Patients in the COVID-19 arm were younger than controls (of 42.8, SD 15.5 vs 46.0, SD 14.0 years; *P*=.02). Patients with COVID-19 were 32% (39/122) female, and the controls were 22% (129/588) female (*P*=.33). There were no other significant differences in baseline characteristics and comorbidities between the 2 arms. Notably, both populations tended to be young and healthy, with most participants having few or no comorbidities. In the group with COVID-19, most of the patients considered (n=112, 92%) were not hospitalized during their COVID-19 diagnosis. Data were collected 171 (SD 114) days after their COVID-19 diagnosis. All baseline characteristics for both COVID-19 and control groups are represented in Table 1.

Table 1. Baseline demographic and clinical characteristics of COVID-19 and control arms.

Characteristics	COVID-19 (n=122)	Control (n=588)	*P* value
Age (years) mean (SD)	41.32 (15.7)	45.99 (14.0)	.001
Gender, n (%)			.33
Male	76 (62)	453 (77)	
Female	46 (38)	135 (23)	
BMI (kg/m^2)	28.7 (8.6)	27.1 (5.7)	.001
Race or ethnicity, n (%)			.36
White	71 (58)	465 (79)	
African American or Black	20 (16.5)	3 (0.5)	
Asian	12 (10)	29 (5)	
Latino or Hispanic	5 (4.5)	41 (7)	
Others	13 (11)	50 (8.5)	
Comorbidity, n (%)			.96
None	88 (72)	506 (86)	
Diabetes	6 (5)	12 (2)	
Immune system deficiencies or HIV	1 (1)	12 (2)	
Heart conditions	4 (3)	12 (2)	
Asthma or chronic lung disease	15 (12)	24 (4)	
Extreme obesity	5 (4)	18 (3)	
Cancer treatment	4 (3)	6 (1)	
Education level, n (%)			.21
Bachelor's degree	27 (22)	247 (42)	
Some college	30 (24)	65 (11)	
Associate degree	16 (13)	41 (7)	
Master's degree	28 (23)	112 (19)	
Doctorate	1 (1)	35 (6)	
Professional	10 (8)	59 (10)	
Others	11 (9)	29 (5)	

Average Follow-up of the 2 Groups Using the Biostrap Device

Controls were followed up for 64 (SD 28) days and patients with long COVID-19 were followed up for 55 (SD 66) days. For the weighted analysis, 37,709 recorded days (103.2 years) were collected for the control group and 7228 recorded days (19.8 years) were collected for patients with COVID-19.

Correlations Between Biometrics and the Different Phases of the Sleeping Cycle

All the correlations between sleep phases and biometrics are summarized in Figure 4.

Total Sleep Cycle

Total sleep time was correlated with RR (r=0.084, P≤.05 for cohort and r=0.119, P≤.01 for controls) and SpO_2 (r=0.076, P≤.05 for cohort and r=0.123, P≤.01 for controls). Total sleep time was not significantly correlated with HR (P>.05) and HRV (P>.05).

Awake Sleep Phase

Significant correlations were found between HR (r=0.109, P≤.01 for cohort and r=0.148, P<.001 for controls), RR (r=–0.201, P<.001 for cohort and r=–0.161, P<.001 for controls), HRV (r=–0.099, P≤.01 for cohort and r=–0.105, P≤.05 for controls), SpO_2 (r=–0.205, P<.001 for cohort and r=–0.162, P<.001 for controls), and awake sleep phase.

Light Sleep Phase

For light sleep phase, only RR (r=0.358, P<.001) and SpO_2 (r=0.249, P<.001) in the COVID-19 group were found to be correlated with the time spent in this phase. There was no significant correlation between light sleep and HR nor between light sleep and HRV (P>.05).

Deep Sleep Phase

As for deep sleep, the time spent in this phase was correlated with HR (r=–0.093, P≤.05 for cohort and r=–0.098, P≤.05 for controls), RR (r=0.121, P≤.01 for cohort and r=0.108, P≤.01 for controls), and SpO_2 (r=0.106, P<.001 for cohort and r=0.094, P≤.01 for controls). However, it did not significantly correlate with HRV (P>.05). As seen in Figure 4, awake sleep significantly correlates with HR in all participants (r=0.121, P<.01), in the control group (r=0.148, P<.001), but not in patients with COVID-19 (r=0.097, P>.05); awake sleep also significantly correlates with RR in all participants (r=–0.205, P<.001), in the control group (r=–0.161, P<.001), and in patients with COVID-19 (r=–0.252, P<.01); awake sleep also correlates with HRV in all participants (r=–0.094, P<.05), in the control group (r=–0.094, P<.05), but not in patients with COVID-19 (r=–0.010, P>.05); awake sleep correlates with SpO_2 in all participants (r=–0.200, P<.001), in the control group (r=–0.162, P<.001), and in patients with COVID-19 (r=–0.187, P<.05). The same interpretation can be drawn from Figure 4 for light sleep, deep sleep, and total sleep.

Comparison of Sleep Cycle in Patients With Lower vs Higher HR

After dividing the cohort into patients with higher HR (50/710 patients, 7%) and lower HR (660/710 patients, 93%), patients with higher HR had more time in awake sleep (65 minutes vs 55 minutes, P=.02) and less time in light (232 minutes vs 258 minutes, P=.001), deep (128 minutes vs 135 minutes, P=.1), and total sleep (425 minutes vs 449 minutes, P=.006; Figure 5). The number of patients and the different results are listed in Table 2.

Figure 5. Summary representation of propensity score matching for age, BMI, and gender. F: Female; M: Male. BMI: body mass index.

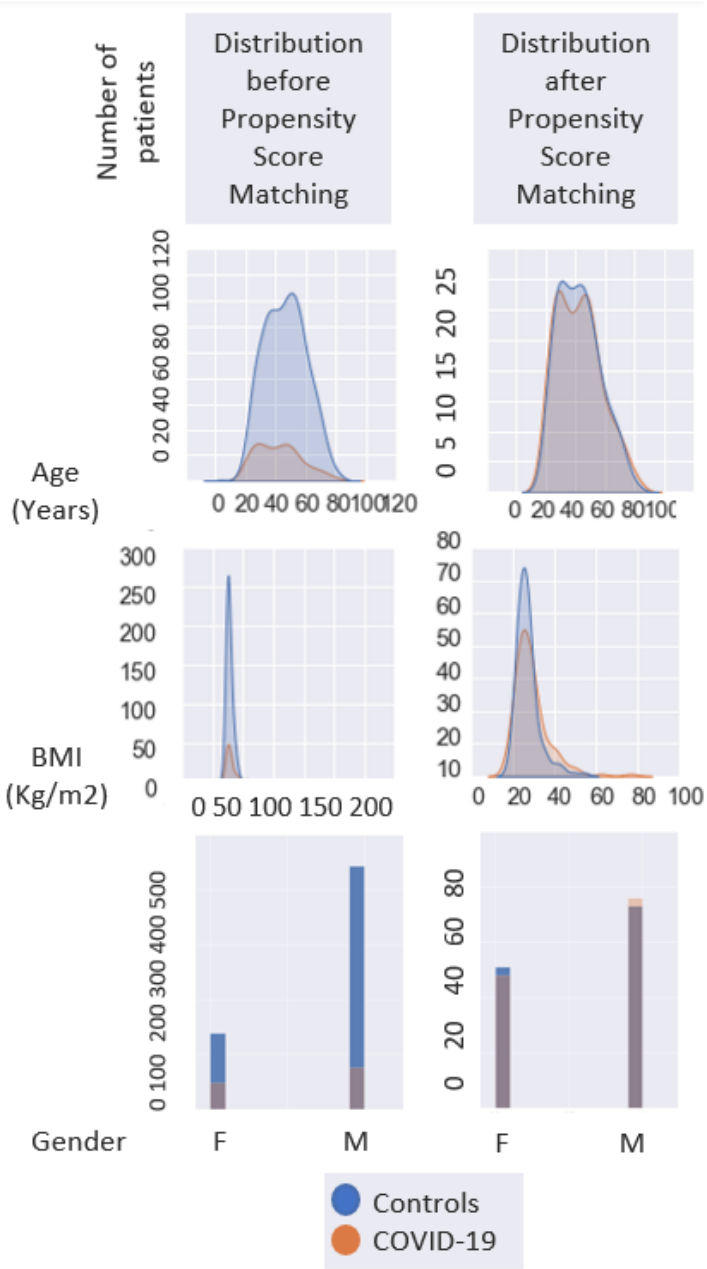


Table 2. Different biometric groups with respective number of patients.

Group	Number of patients	Awake sleep phase (min)[a]	Light sleep phase (min)[b]	Deep sleep phase (min)[c]	Total sleep phase (min)[d]
Higher HR[e] (>80 beats/min)	50	65	232	128	425
Lower HR (<80 beats/min)	660	55	258	135	449
Higher RR[f] (>20 respirations per minute)	0	—[g]	—	—	—
Lower RR (<20 respirations per minute)	710	—	—	—	—
Higher HRV[h] (>20ms)	683	—	—	—	—
Lower HRV (<20ms)	27	—	—	—	—

[a]*P*=.02.

[b]*P*=.001.

[c]*P*=.1.

[d]*P*=.006.

[e]HR: heart rate.

[f]RR: respiratory rate.

[g]Statistical analysis was not performed to assess the differences between these groups because of the low number of patients in Higher RR and Lower HRV groups.

[h]HRV: heart rate variability.

Comparison of Sleep Length Between Patients With COVID-19 and Controls

Unweighted Analysis

In the unweighted analysis, patients with long COVID-19 had less total sleep time when compared to controls (433, SD 85 vs 450, SD 68; *P*<.001).

Weighted Analysis

In the weighted analysis, patients with long COVID-19 had statistically but not clinically significant increased total sleep time when compared to control group (451.4, SD 65 vs 451.7, SD 87 minutes, *P*<.001).

Propensity Match Analysis

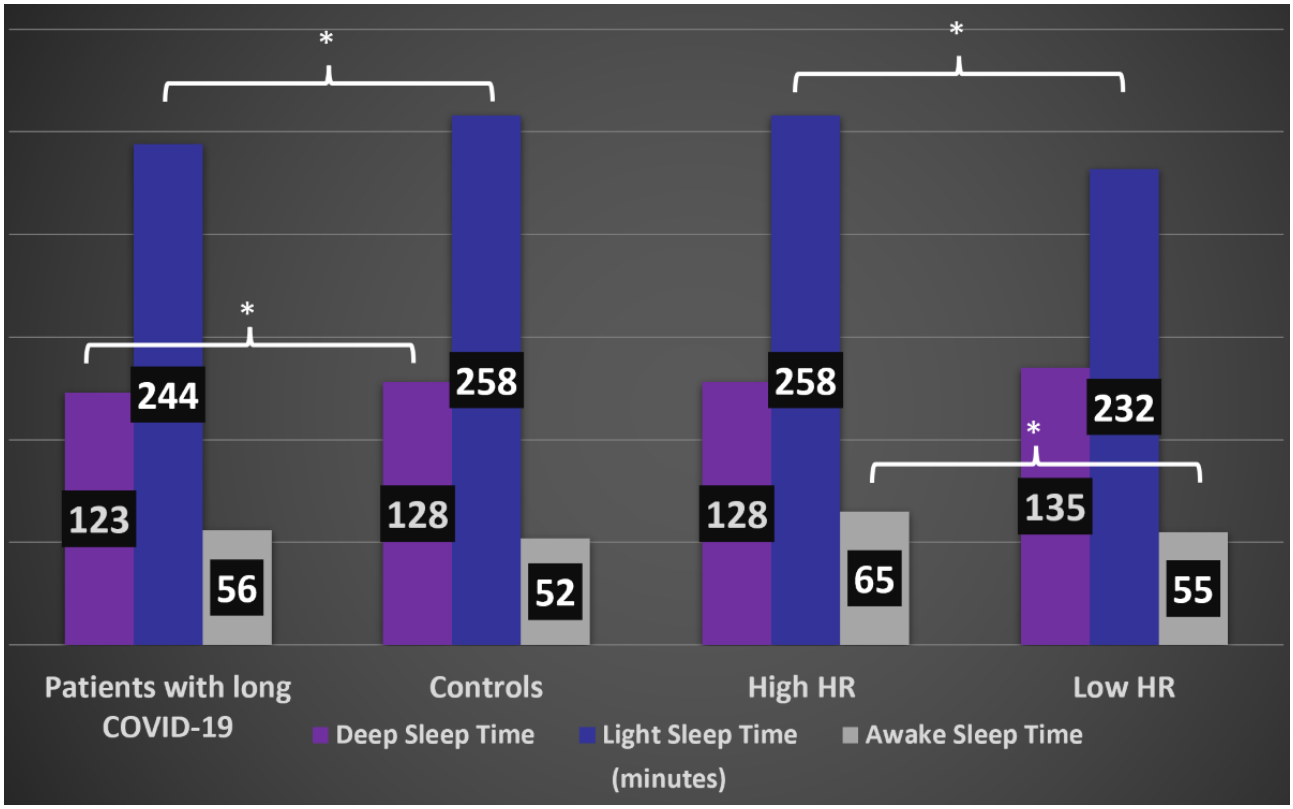
After performing a propensity match analysis, 122 patients with COVID-19 were compared to 122 matched controls. Total sleep time was found to be decreased in the group with COVID-19 compared to controls (433, SD 85 vs 450, SD 68; *P*=.004). A schematic representation of the data distribution before and after propensity score matching was shown for better visualization (Figure 5).

Comparison of Sleep Cycle Phases Between Patients With COVID-19 and Controls

Unweighted Analysis

In the unweighted analysis, there was no statistical awake sleep time difference between the 2 groups (52, SD 32 for controls vs 56, SD 31 for cohort; *P*=.4). However, patients with long COVID-19 had decreased light sleep (244, SD 67 vs 258, SD 67; *P*=.003) and decreased deep sleep (123, SD 66 vs 128, SD 58; *P*=.02; Figure 6).

Figure 6. Difference in weighted sleep phases between different groups. High heart rate (HR): >80 beats per minute. Low HR: <80 beats per minute; *P<.05.



Weighted Analysis

When comparing weighted different sleep phases, patients with long COVID-19 had increased awake sleep time (62, SD 25 vs 52, SD 32; P<.001) decreased light sleep time (251, SD 82 vs 260, SD 64; P<.001), and decreased deep sleep (126, SD 71 vs 131, SD 59; P<.001).

Propensity Score Matching

After matching the 2 groups for age, sex, and BMI, patients with COVID-19 have decreased deep sleep when compared to controls (123, SD 66 vs 128, SD 57; P<.004). However, the differences in light sleep (244, SD 67 vs 259, SD 67; P=.39) and awake sleep (56, SD 31 vs 57, SD 32; P=.71) were no longer significant.

Discussion

Overview

In our study, we report 2 major findings. First, increased total sleep time and time spent in deep sleep were associated with increased RR and SpO_2, and decreased HR in the full cohort. Second, the group with long COVID-19 had altered sleep architecture characterized by decreased total and deep sleep times when compared to matched controls.

Association Between Biometrics and Sleep Phases

Decreased oxygen saturation during sleep can be due to different pathologies and has the potential to inflict significant negative physiological and psychological consequences [27]. In our cohort, decreased RR and SpO_2 were associated with increased time in awake sleep phase. Moreover, increased RR and SpO_2 were associated with increased total sleep time and deep sleep. Our findings suggest difficulty transitioning into deep sleep of patients with decreased respiratory function and thus the need for good oxygenation and respiratory function to maintain a physiological sleep cycle. This is in line with the increased sleeping disturbances noticed in patients with severe COVID-19. Huang et al [28] showed that the risk of severe infection was 6 to 8 times more associated with decreased sleep status and reduced sleeping hours [28]. Additionally, the reduction in average daily sleep time significantly increased the likelihood of infection severity, stressing on the intertwined relationship between sleep and respiratory function [28]. However, the results reported in this study were extracted from self-reported questionnaires in comparison to our quantitative results. Moreover, increased RR and SpO_2 in the group with COVID-19 was correlated with increased time spent in light sleep. Light sleep, which is one of the different phases of NREM sleep, was found to have an important role in memory formation and consolidation as well as in motor skill speed and performance [29,30]. Therefore, the association of optimal respiratory function during sleep and improved sleep quality may improve activities of daily living and quality of life in addition to immunity and response to infections.

As for the autonomous system, increased HR and decreased HRV were correlated with increased time in awake phase, whereas decreased HR was correlated with increased time in deep sleep. Increased HR during sleep was associated with increased cardiovascular comorbidities [31]. This was widely studied in night-shift workers, who had misalignment between the endogenous circadian clock and the sleep schedule, leading

to increased cardiovascular events [32]. Our data showing that participants with increased HR have increased time in awake sleep and less time in deep sleep might indicate a difficulty in transitioning from light to deep sleep among patients with increased HR. Disturbances in the configuration of these 2 systems may lead to adverse repercussions and clinical outcomes.

The key benefit of continuous monitoring with wearables lies in the capability to detect these vulnerable populations who may have early sleep or biometric disturbances. The collection of real-time data from wearables can allow the physician to manage patients at a very early stage. By combining data from biometrics and sleep phases, physicians will be able to have an overview on patients' autonomic system activity. These findings, sometimes subclinical, will be useful as a significant decision support tool for physicians to employ preventative and personalized medicine even before diagnosing the problem.

Long COVID-19's Effect on Sleep

Previous studies have shown increased psychological disturbances in addition to the physical component associated with COVID-19 infection [22,33]. In fact, decreased sleep quality and insomnia problems have been reported during the pandemic. However, most sleep-related studies focused primarily on health care workers rather than the general population [34,35]. For example, Zhang et al [35] found that almost one-third of health care workers had insomnia symptoms during the pandemic, and that the related factors included education level, isolation environment, and psychological stressors [35]. However, most of these studies were qualitative and survey-based rather than quantitative [22,23]. Thus, there was a need for a quantitative approach to assess long COVID-19's effect on sleep.

In our study, participants in the group with long COVID-19 had increased awake sleep time and decreased light and deep sleep time. During sleep, the body secures restorative functions related to immunity [36], the cardiovascular system [37], and metabolic functions [38]. Alterations in non-REM sleep phases may therefore predispose health-related problems. In addition, altered sleep architecture was shown to increase stress levels by increasing stress hormones [39-42]. These findings especially after recovering from the infection support the fact that COVID-19 may present with long-standing symptoms such as autonomic and neurologic disturbances. This is in alignment with what is called "Long COVID-19" syndrome or "COVID-19 Brain fog," which is characterized by fatigue, difficulty concentrating, and sleep disorders even after the acute infection [43].

Limitations

Our study has several limitations. First, this is a single-center study, limiting the reproducibility of the results among a wider population. Second, baseline data regarding physiological state of participants with COVID-19 is not available as they did not have the device before COVID-19 infection. Third, the PPG technology used was not developed to accurately characterize REM sleep, and thus, REM sleep has been omitted from the analysis. Finally, the controls were recruited based on a patient-reported survey, and therefore they might have had COVID-19 without knowing.

Conclusion

Study participants with improved cardiovascular and respiratory functions had better sleep architecture. Moreover, patients who were diagnosed with COVID-19, including young and healthy patients, demonstrated altered sleep architecture when compared to matched controls. The sleeping data of patients with COVID-19 were characterized by decreased total sleep and deep sleep times. Future studies should evaluate the physical and psychological impact of sleep disturbance among patients with long COVID-19.

Acknowledgments

We thank the Brown foundation for providing support and funding for this study.

Data Availability

The data that support the findings of this study are available from the corresponding author, N Marrouche, upon reasonable request.

Conflicts of Interest

N Marrouche reports receiving grant support from Abbott, Medtronic, Biosense Webster, Boston Scientific and consulting fees from Preventice, Biosense Webster, Atricure lectures: Biotronik, Bristol Myers Squibb and Biosense Webster. MM, CHL, AHEH, CN, CP, N Makan, LD, YZ, NC, DLL, and TA declare no conflicts of interest.

References

1. Ackermann M, Verleden SE, Kuehnel M, Haverich A, Welte T, Laenger F, et al. Pulmonary Vascular Endothelialitis, Thrombosis, and Angiogenesis in Covid-19. N Engl J Med 2020 Jul 09;383(2):120-128. [doi: 10.1056/nejmoa2015432]
2. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. The Lancet 2020 Mar;395(10227):912-920. [doi: 10.1016/s0140-6736(20)30460-8]

3. Öngür D, Perlis R, Goff D. Psychiatry and COVID-19. JAMA 2020 Sep 22;324(12):1149-1150. [doi: 10.1001/jama.2020.14294] [Medline: 32960247]
4. Khan S, Siddique R, Li H, Ali A, Shereen MA, Bashir N, et al. Impact of coronavirus outbreak on psychological health. J Glob Health 2020 Jun;10(1):010331 [FREE Full text] [doi: 10.7189/jogh.10.010331] [Medline: 32355556]
5. Gunasekeran DV, Tseng RMWW, 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]
6. Patel AK, Ready V, Araujo JF. Physiology, Sleep Stages. National Library of Medicine. 2022. URL: https://www.ncbi.nlm.nih.gov/books/NBK526132/ [accessed 2022-06-28]
7. Wolpert EA. A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects. Arch Gen Psychiatry 1969 Feb 01;20(2):246. [doi: 10.1001/archpsyc.1969.01740140118016]
8. Hirshkowitz M. Assessing sleep wearables and in-bedroom devices: CTA standards work. SlidePlayer. 2019. URL: https://slideplayer.com/slide/13756482/ [accessed 2022-06-28]
9. Nanjapan V, Liang H, Wang W, Man KL. Chapter 1 - Big Data: A Classification of Acquisition and Generation Methods. In: Big Data Analytics for Sensor-Network Collected Intelligence. Cambridge, MA, US: Academic Press; 2017:3-20.
10. Spierer DK, Rosen Z, Litman LL, Fujii K. Validation of photoplethysmography as a method to detect heart rate during rest and exercise. J Med Eng Technol 2015 Jun 26;39(5):264-271. [doi: 10.3109/03091902.2015.1047536] [Medline: 26112379]
11. Trinder J, Kleiman J, Carrington M, Smith S, Breen S, Tan N, et al. Autonomic activity during human sleep as a function of time and sleep stage. J Sleep Res 2001 Dec 13;10(4):253-264 [FREE Full text] [doi: 10.1046/j.1365-2869.2001.00263.x] [Medline: 11903855]
12. Kryger M, Roth T, Dement WC. Principles and Practice of Sleep Medicine, Sixth Edition. Philadelphia, PA, US: Elsevier; 2017.
13. Redmond SJ, de Chazal P, O'Brien C, Ryan S, McNicholas WT, Heneghan C. Sleep staging using cardiorespiratory signals. Somnologie 2007 Oct 11;11(4):245-256. [doi: 10.1007/s11818-007-0314-8]
14. Domingues A, Paiva T, Sanches JM. Hypnogram and Sleep Parameter Computation From Activity and Cardiovascular Data. IEEE Trans. Biomed. Eng 2014 Jun;61(6):1711-1719. [doi: 10.1109/tbme.2014.2301462]
15. Willemen T, Van Deun D, Verhaert V, Vandekerckhove M, Exadaktylos V, Verbraecken J, et al. An Evaluation of Cardiorespiratory and Movement Features With Respect to Sleep-Stage Classification. IEEE J. Biomed. Health Inform 2014 Mar;18(2):661-669. [doi: 10.1109/jbhi.2013.2276083]
16. Fonseca P, Long X, Radha M, Haakma R, Aarts RM, Rolink J. Sleep stage classification with ECG and respiratory effort. Physiol Meas 2015 Oct;36(10):2027-2040. [doi: 10.1088/0967-3334/36/10/2027] [Medline: 26289580]
17. 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]
18. Altmann DM, Boyton RJ. Decoding the unknowns in long covid. BMJ 2021 Feb 04;372:n132. [doi: 10.1136/bmj.n132] [Medline: 33541867]
19. Brodin P. Immune determinants of COVID-19 disease presentation and severity. Nat Med 2021 Jan 13;27(1):28-33. [doi: 10.1038/s41591-020-01202-8] [Medline: 33442016]
20. Targa ADS, Benítez ID, Moncusí-Moix A, Arguimbau M, de Batlle J, Dalmases M, et al. Decrease in sleep quality during COVID-19 outbreak. Sleep Breath 2021 Jun 29;25(2):1055-1061 [FREE Full text] [doi: 10.1007/s11325-020-02202-1] [Medline: 32989674]
21. Xiao H, Zhang Y, Kong D, Li S, Yang N. Social Capital and Sleep Quality in Individuals Who Self-Isolated for 14 Days During the Coronavirus Disease 2019 (COVID-19) Outbreak in January 2020 in China. Med Sci Monit 2020 Mar 12;26:e923921. [doi: 10.12659/msm.923921]
22. 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 Res 2020 Jun;288:112954 [FREE Full text] [doi: 10.1016/j.psychres.2020.112954] [Medline: 32325383]
23. Vindegaard N, Benros ME. COVID-19 pandemic and mental health consequences: Systematic review of the current evidence. Brain Behav Immun 2020 Oct;89:531-542 [FREE Full text] [doi: 10.1016/j.bbi.2020.05.048] [Medline: 32485289]
24. Dur O, Rhoades C, Ng MS, Elsayed R, van Mourik R, Majmudar MD. Design Rationale and Performance Evaluation of the Wavelet Health Wristband: Benchtop Validation of a Wrist-Worn Physiological Signal Recorder. JMIR Mhealth Uhealth 2018 Oct 16;6(10):e11040 [FREE Full text] [doi: 10.2196/11040] [Medline: 30327288]
25. Gielen W, Longoria KA, van Mourik RA. Two cases of COVID-19 monitored by a wearable biosensor-a case report. Mhealth 2021 Oct;7:62-62 [FREE Full text] [doi: 10.21037/mhealth-20-134] [Medline: 34805393]
26. Nanchen D. Resting heart rate: what is normal? Heart 2018 Jul 30;104(13):1048-1049. [doi: 10.1136/heartjnl-2017-312731] [Medline: 29382691]
27. Szabó K, Ihász F. [The effect of reduced oxygen saturation during sleep on depression]. Orv Hetil 2019 May;160(20):780-783. [doi: 10.1556/650.2019.31401] [Medline: 31081357]
28. Huang B, Niu Y, Zhao W, Bao P, Li D. Reduced Sleep in the Week Prior to Diagnosis of COVID-19 is Associated with the Severity of COVID-19. NSS 2020 Nov;Volume 12:999-1007. [doi: 10.2147/nss.s263488]

29. Hong J, Ha GE, Kwak H, Lee Y, Jeong H, Suh P, et al. Destabilization of light NREM sleep by thalamic PLCβ4 deletion impairs sleep-dependent memory consolidation. Sci Rep 2020 Jun 01;10(1):8813 [FREE Full text] [doi: 10.1038/s41598-020-64377-7] [Medline: 32483199]
30. Walker MP, Brakefield T, Morgan A, Hobson J, Stickgold R. Practice with Sleep Makes Perfect. Neuron 2002 Jul;35(1):205-211. [doi: 10.1016/s0896-6273(02)00746-8]
31. Boudreau P, Yeh W, Dumont GA, Boivin DB. Circadian variation of heart rate variability across sleep stages. Sleep 2013 Dec 01;36(12):1919-1928 [FREE Full text] [doi: 10.5665/sleep.3230] [Medline: 24293767]
32. Frost P, Kolstad HA, Bonde JP. Shift work and the risk of ischemic heart disease - a systematic review of the epidemiologic evidence. Scand J Work Environ Health 2009 May 22;35(3):163-179 [FREE Full text] [doi: 10.5271/sjweh.1319] [Medline: 19387517]
33. Salari N, Hosseinian-Far A, Jalali R, Vaisi-Raygani A, Rasoulpoor S, Mohammadi M, et al. Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: a systematic review and meta-analysis. Global Health 2020 Jul 06;16(1):57 [FREE Full text] [doi: 10.1186/s12992-020-00589-w] [Medline: 32631403]
34. Bao Y, Sun Y, Meng S, Shi J, Lu L. 2019-nCoV epidemic: address mental health care to empower society. The Lancet 2020 Feb;395(10224):e37-e38. [doi: 10.1016/s0140-6736(20)30309-3]
35. Zhang C, Yang L, Liu S, Ma S, Wang Y, Cai Z, et al. Survey of Insomnia and Related Social Psychological Factors Among Medical Staff Involved in the 2019 Novel Coronavirus Disease Outbreak. Front Psychiatry 2020 Apr 14;11:306 [FREE Full text] [doi: 10.3389/fpsyt.2020.00306] [Medline: 32346373]
36. Besedovsky L, Lange T, Born J. Sleep and immune function. Pflugers Arch 2012 Jan 10;463(1):121-137 [FREE Full text] [doi: 10.1007/s00424-011-1044-0] [Medline: 22071480]
37. Jackson CL, Redline S, Emmons KM. Sleep as a potential fundamental contributor to disparities in cardiovascular health. Annu Rev Public Health 2015 Mar 18;36(1):417-440 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122838] [Medline: 25785893]
38. Van Cauter E, Holmback U, Knutson K, Leproult R, Miller A, Nedeltcheva A, et al. Impact of sleep and sleep loss on neuroendocrine and metabolic function. Horm Res 2007 Feb 15;67 Suppl 1(Suppl. 1):2-9. [doi: 10.1159/000097543] [Medline: 17308390]
39. Sanford L, Suchecki D, Meerlo P. Stress, arousal, and sleep. Curr Top Behav Neurosci 2015;25:379-410. [doi: 10.1007/7854_2014_314] [Medline: 24852799]
40. Minkel J, Moreta M, Muto J, Htaik O, Jones C, Basner M, et al. Sleep deprivation potentiates HPA axis stress reactivity in healthy adults. Health Psychol 2014 Nov;33(11):1430-1434. [doi: 10.1037/a0034219] [Medline: 24818608]
41. Leproult R. Sleep loss results in an elevation of cortisol levels the next evening. Sleep 1997;20(10):865-870. [doi: 10.1093/sleep/20.10.865]
42. Redwine L. Effects of Sleep and Sleep Deprivation on Interleukin-6, Growth Hormone, Cortisol, and Melatonin Levels in Humans. Journal of Clinical Endocrinology & Metabolism 2000 Oct 01;85(10):3597-3603. [doi: 10.1210/jc.85.10.3597]
43. 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]

Abbreviations

HR: heart rate
HRV: heart rate variability
NREM: nonrapid eye movement
PPG: photoplethysmography
REM: rapid eye movement
RR: respiratory rate
SpO_2: oxygen saturation
WEAICOR: Wearables to Investigate the Long Term Cardiovascular and Behavioral Impacts of COVID-19

Edited by C Basch; submitted 15.03.22; peer-reviewed by A Joseph, M Tummalacherla, K Uludag; comments to author 14.04.22; revised version received 10.05.22; accepted 20.06.22; published 05.07.22.

Please cite as:
Mekhael M, Lim CH, El Hajjar AH, Noujaim C, Pottle C, Makan N, Dagher L, Zhang Y, Chouman N, Li DL, Ayoub T, Marrouche N
Studying the Effect of Long COVID-19 Infection on Sleep Quality Using Wearable Health Devices: Observational Study
J Med Internet Res 2022;24(7):e38000
URL: https://www.jmir.org/2022/7/e38000
doi:10.2196/38000
PMID:35731968

Original Paper

Deep Denoising of Raw Biomedical Knowledge Graph From COVID-19 Literature, LitCovid, and Pubtator: Framework Development and Validation

Chao Jiang[1], BS; Victoria Ngo[2,3,4], PhD; Richard Chapman[1], PhD; Yue Yu[5], PhD; Hongfang Liu[5], PhD; Guoqian Jiang[5], MD, PhD; Nansu Zong[5], PhD

[1]Department of Computer Science and Software Engineering, Auburn University, Auburn, AL, United States
[2]Center for Innovation to Implementation, VA Palo Alto Health Care System, Sacramento, CA, United States
[3]Stanford Health Policy, Stanford School of Medicine, Stanford University, Stanford, CA, United States
[4]Freeman Spogli Institute for International Studies, Stanford University, Stanford, CA, United States
[5]Department of Artificial Intelligence and Informatics Research, Mayo Clinic, Rochester, MN, United States

Corresponding Author:
Nansu Zong, PhD
Department of Artificial Intelligence and Informatics Research
Mayo Clinic
200 First St SW
Rochester, MN, 55905
United States
Phone: 1 507 284 2511
Email: Zong.Nansu@mayo.edu

Abstract

Background: Multiple types of biomedical associations of knowledge graphs, including COVID-19–related ones, are constructed based on co-occurring biomedical entities retrieved from recent literature. However, the applications derived from these raw graphs (eg, association predictions among genes, drugs, and diseases) have a high probability of false-positive predictions as co-occurrences in the literature do not always mean there is a true biomedical association between two entities.

Objective: Data quality plays an important role in training deep neural network models; however, most of the current work in this area has been focused on improving a model's performance with the assumption that the preprocessed data are clean. Here, we studied how to remove noise from raw knowledge graphs with limited labeled information.

Methods: The proposed framework used generative-based deep neural networks to generate a graph that can distinguish the unknown associations in the raw training graph. Two generative adversarial network models, NetGAN and Cross-Entropy Low-rank Logits (CELL), were adopted for the edge classification (ie, link prediction), leveraging unlabeled link information based on a real knowledge graph built from LitCovid and Pubtator.

Results: The performance of link prediction, especially in the extreme case of training data versus test data at a ratio of 1:9, demonstrated that the proposed method still achieved favorable results (area under the receiver operating characteristic curve >0.8 for the synthetic data set and 0.7 for the real data set), despite the limited amount of testing data available.

Conclusions: Our preliminary findings showed the proposed framework achieved promising results for removing noise during data preprocessing of the biomedical knowledge graph, potentially improving the performance of downstream applications by providing cleaner data.

(J Med Internet Res 2022;24(7):e38584) doi:10.2196/38584

KEYWORDS

adversarial generative network; knowledge graph; deep denoising; machine learning; COVID-19; biomedical; neural network; network model; training data

Introduction

The effects of the COVID-19 pandemic linger in 2022—it affected over 11.6 million people globally in the past year, and accounted for >2.5 million deaths in more than 220 countries [1]. With the continuous accumulation of peer-reviewed publications on the topic, a literature hub serves as a means to track the most up-to-date scientific information about the virus [2]—encompassing research on the treatment, diagnosis, and prevention of COVID-19. A knowledge base built upon the integration of biomedical entities from such a literature hub would provide tremendous value in the exploration of explicit or implicit associations among diverse biomedical entities as investigators attempt to answer clinical questions related to COVID-19. A number of recently published journal articles have included graph-based analysis of COVID-19 data sets [3]. For example, Groza [4] analyzed how a semantically annotated data set would be helpful in detecting and preventing potentially harmful misinformation regarding the spread of COVID-19 based on CORD-19-on-FHIR (a linked data version of the COVID-19 Open Research Dataset [CORD-19] data represented in FHIR RDF by mining the CORD-19 data set and adding semantic annotations) [5].

Most knowledge graphs constructed for COVID-19 are currently based on the co-occurring biomedical entities reported in recent literature. A knowledge graph of co-occurring concepts, such as the one created by Oniani et al [6], can help researchers find associations among genes, drugs, and diseases related to COVID-19. Using knowledge graphs with heterogeneous biomedical associations (eg, gene-drug, disease-drug, drug–side effect) in these types of applications, however, results in a high probability of false-positive predictions because co-occurrence in literature does not always mean there is a true biomedical association between the two entities. These co-occurrence edges are therefore considered "noise" due to their untrue associations. For example, the term "glucose" may co-occur with the term "yellow fever," but there is no real medical association between the two terms. Noise removal can be beneficial for downstream applications, such as link prediction [7], representation learning [8,9], and node classification [10].

The manual processes of cleaning data and removing noise are resource intensive. Therefore, an automated denoising method is ideal in facilitating the curation of knowledge graphs. Existing methods for denoising knowledge graphs can be divided into two groups: internal and external [11]. For the internal method, the predefined semantics or rules [12] are used for nonnumerical data. Outlier detection [13] removes noise by modeling true data as a distribution for numerical data. As for external methods, a pretrained graph neural network integrates heterogeneous data sources [14] to not only improve the performance of link prediction but also reduce the training time of the existing graph neural network model. In this paper, our methodology can be categorized as an internal method where data augmentation with a generative adversarial network (GAN) removes noise. GAN has been widely applied in medical imaging process [15] to denoise computed tomography images based on GAN with Wasserstein distance and perceptual similarity. Zhou et al [16] previously showed improvement of ultrasonic image quality and noise reduction caused device limitations through the construction of a two-stage GAN. Other than the application of generating images, GAN has mainly been used for generating discrete medical data to contribute to the scenario of diagnosis of a disease with few labels [17] or unbalanced classification [18]. To the best knowledge of the authors, our study is the first study that uses GAN to denoise a biomedical knowledge graph.

Here, we propose a framework that generates a similar graph from a raw knowledge graph to distinguish the true and false edges of association based on generative-based deep neural networks. Two recent generative-based models, Cross-Entropy Low-rank Logits (CELL) [19] and a generative-based graph method (NetGAN) [20], have been adopted as a component to remove noise and retain true associations within two data sets: (1) a synthetic data set generated from CORA-ML [21] with the same preprocessing as in NetGAN [20]; and (2) a real data set constructed from CORD-19-on-FHIR data sets with heterogeneous biomedical associations (ie, chemical-disease, gene-disease, gene-chemical associations) [5]. Our study shows the proposed method achieved promising results in the classification tasks for separating the true and negative edges.

Methods

Problem Definition

Given a network G(V, E), where *V* stands for a set of vertices (ie, biomedical concepts in the literature) and *E* represents the edges among two vertices (ie, the co-occurrence of two concepts), two kinds of edges exist, which are denoted as *L* (known true associations) and *U* (unknown true associations). We note that if no edge exists between two vertices, this will be considered a false association. The aim is to find the true associations among *U* (ie, denoise *U*). Specifically, a proposed method should have the capacity to determine whether unknown true associations from *U* are true associations or false.

Framework

Overview

As this problem could be considered a classification of an unknown edge with a small number of known true associations and a large number of unknown true associations, we defined this classification problem as few-shot learning [22]. We proposed a framework that used generative-based deep neural networks (eg, NetGAN and CELL) to denoise the unknown true associations in *U* based on similar networks generated. This framework was divided into 3 parts. We first briefly describe the GAN-based denoising graph adopted following the development of the framework, followed by an introduction of data preparation, which involved two strategies: (1) synthetic data generation and (2) real data set collection and annotation. A comprehensive design of our experiments was then conducted to verify our assumptions.

Denoising Based on Generative-Based Deep Neural Networks

We adopted NetGAN to generate a new network that would be used to distinguish the unknown associations in the raw training

data (ie, graph). To achieve this, we randomly sampled walks from the raw graph consisting of unlabeled edges and trained a generator to learn the walks sampled and a discriminator on how to separate a real walk from a fake one. After achieving equilibrium among the discriminator and generator, the random walk sample from the generator was used for filtering the unreal edge in the raw graph. As determined in previous work by other researchers [19], sampling enough random walks was sufficient to reconstruct the graph. Both the generator and discriminator used the long short-term memory (LSTM) architecture [23] and were trained with the Wasserstein loss [24]. The generator *G* generated large numbers of random walks (node sequence) of fixed length. The discriminator *D* distinguished the sequence of the nodes sampled from *G* and *x* that were sampled from the real graph (including unlabeled associations) with randomly started nodes. *D* and *G* played the following minimax game with the value function *V(D, G)*:

Finally, the *D* generated a similar authentic graph network that could not be distinguished by the discriminator *G*.

To generate the probability of the edges, CELL approximated it with a score matrix *S*, which was computed by  where *n* is the number of random walks, *T* is each length of a random walk, and diag(π) is the stationary distribution matrix. *P* is a transition matrix that approximates the unbiased random walk used in NetGAN. *P* can be low-rank approximated by *W*, which is the logit transition matrix and is solved by the objective function as:

where *A* is the adjacency matrix and s.t. rank(W) ≤ *H*. In practice, we further adapted node2vec [25] for the random sampling process in NetGAN and constrained the edge generation length with *k* in the above loss function in CELL.

Data Preparation

Overview

We generated two data sets for this study: (1) a synthetic data set based on CORA-ML, and (2) a real data set extracted from CORD-19-on-FHIR data sets [5]. First, we defined two types of associations: labeled associations denoted as (*L*) (red colored) and unlabeled associations represented as (*U*) (green colored) (Figure 1), based on two types of association. This was used to construct our training and test graphs. The training graph consisted of both the labeled (*L*) and unlabeled (*U*) associations, while there were only labeled (*L*) associations in the test graph, as we need the ground truth for evaluating the performance of our proposed methods. The histogram of each data set is given below, where Figure 2A is the synthetic data set. This does not include the false associations added in our subsequent experiments. Figure 2B shows the histogram of degree distribution in the real data set.

Figure 1. Overview of our investigation process. GAN: generative adversarial network; ROC-AUC: area under the receiver operating characteristic curve.

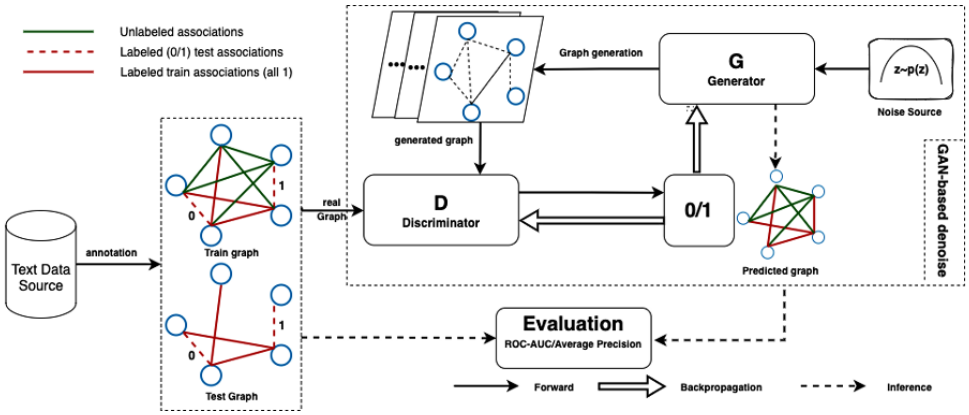
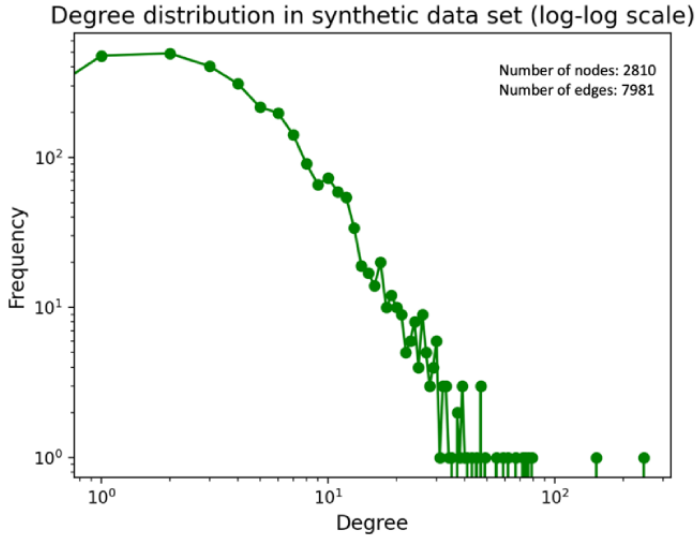
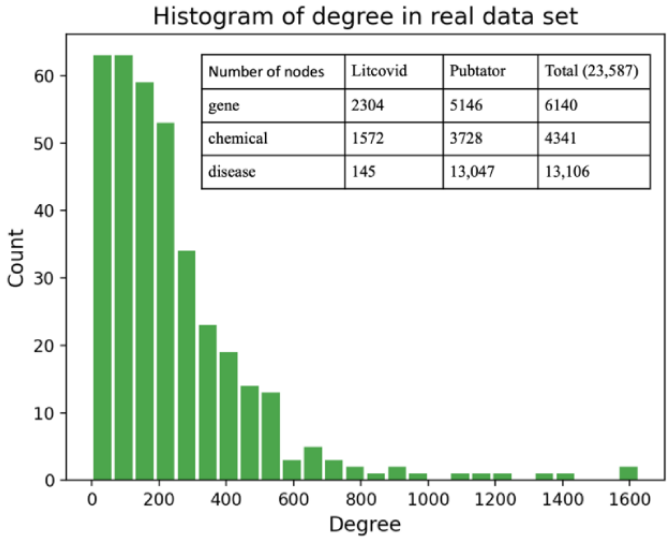


Figure 2. Histogram of degree distribution in the synthetic data set and real data set.



(A) Synthetic data set



Number of nodes	Litcovid	Pubtator	Total (23,587)
gene	2304	5146	6140
chemical	1572	3728	4341
disease	145	13,047	13,106

(B) Degree

Synthetic Data Set

The synthetic data set was generated based on CORA-ML with the same preprocessing work as NetGAN [20]; we chose the largest connected component in the graph. The final total number of nodes and edges is shown in the top right corner of Figure 2A. To test our proposed methods, we took as ground truth the existing edges as true associations and the nonexistence edges as false associations. Detailed synthetic data processes can be found in Figure S1 in Multimedia Appendix 1.

Real Data Set

From CORD-19-on-FHIR [5], we used two annotated networks (LitCovid [26], Pubtator [27]), extracting the COVID-19–related terms in our SPARQL query with 3 types of biomedical concepts (ie, gene, chemical, mutation/disease). After merging identical IDs from both LitCovid and Pubtator, we were able to obtain a new data set with a total of 23,578 nodes (Figure 2B). Finally, we randomly chose a proportional number of edges with a total of 500 associations (ie, chemical-disease, gene-disease, gene-chemical associations) from a total of 288,270 edges in the whole graph and manually labeled them as our labeled data set. Detailed data preprocessing and degree distribution for each type of association can be found in Supplementary 2 in Multimedia Appendix 1.

Experiment Design

Overview

We conducted experiments on both a synthetic data set (ie, CORA-ML) and a real data set extracted from CORD-19-on-FHIR to investigate the capability of our proposed methods of incorporating unlabeled information for improving the link prediction performance despite limited annotation. We analyzed the performance of our models with multiple tasks based on two types of ratios to mimic the percentages of noise and annotation during the data curation: (1) noise ratio (NR), the percentage of true and false associations in the unlabeled edges; and (2) annotation ratio (AR), the percentage of training and testing associations in the labeled edges.

Task 1: Test of AR Over the Synthetic Data Set

We wanted to understand how many annotations were needed during the data curation for our proposed method to predict the true and false associations. We set a fixed NR and evaluated the performance of the tested method in two cases: one included the unlabeled data (ie, training set = labeled true and false associations + unlabeled associations), and the other did not include the unlabeled data (ie, training set = labeled true and false associations). The unlabeled associations were taken as true associations for training. In the experiment, we tested the performances based on different AR to mimic the percentage of the annotations already completed during the data curation. In practice, the AR varied from 1:9 to 9:1. For each ratio, we repeated the test 10 times with a random sampling of the training and testing sets to get the average results.

Task 2: Test of NR Over the Synthetic Data Set

In this task, we wanted to understand how many false associations were deemed as true associations in unlabeled data for training because it affected the prediction performance of the proposed method. We wanted to see whether the proposed method was robust enough to learn useful information for prediction, especially from unlabeled edges with more noise. With a fixed AR of 1:1, we tested the proposed method when there were more false edges than real edges in the unlabeled data. In practice, the NR varied from 1:1 to 1:9.

Task 3: Test Over the Real Data Set

After the same training of annotation, two of the authors (CJ and YY) manually labeled 500 of the 288,270 edges to simulate an extreme use case for data curation, and another author (VN) verified the annotation by random sampling the edges. Among the 500 edges, the 3 types consisted of chemical-disease, gene-chemical, and gene-disease. Each edge was marked as true, false, and unknown. In practice, the annotations for gene-chemical were excluded and marked as unknown in the final evaluation after the authors had a discussion and reached a consensus that those annotations were conducted without enough confidence. Thus, in our final result report of the receiver operating characteristic curve, we only considered 2 types of associations: chemical-disease and gene-disease.

Setting and Evaluation Metrics

For each proposed method (ie, NetGAN and CELL), a grid search strategy was adapted for obtaining the best hyperparameters. In our experiment, we defined the search range by referencing the original settings in the articles. For NetGAN, the parameter ranges for the grid search are specified as walk $p = \{0.01, 0.1, 1, 10, 100\}$ and $q = \{0.01, 0.1, 1, 10, 100\}$. For CELL, the parameter ranges are specified as rank $H = \{9, 20\}$, learning rate $lr = \{0.01, 0.05, 0.1\}$, and weight decay $weight_{decay} = \{1e-5, 1e-6, 1e-7\}$. In practice, the origin NetGAN was obtained from [28], and the origin CELL was obtained from [29].

In the evaluation step, we chose the area under the receiver operating characteristic curve (AUC ROC) and average precision (AP) as the metrics of link prediction for our proposed methods in both synthetic and real data sets. In the implementation, both AUC ROC and AP scores were calculated by scikit-learn [30]. The visualization of predicted results in our real data set was a plot made with Cytoscape [31], an open-source software platform for visualizing complex networks and integrating these with any type of attribute data.

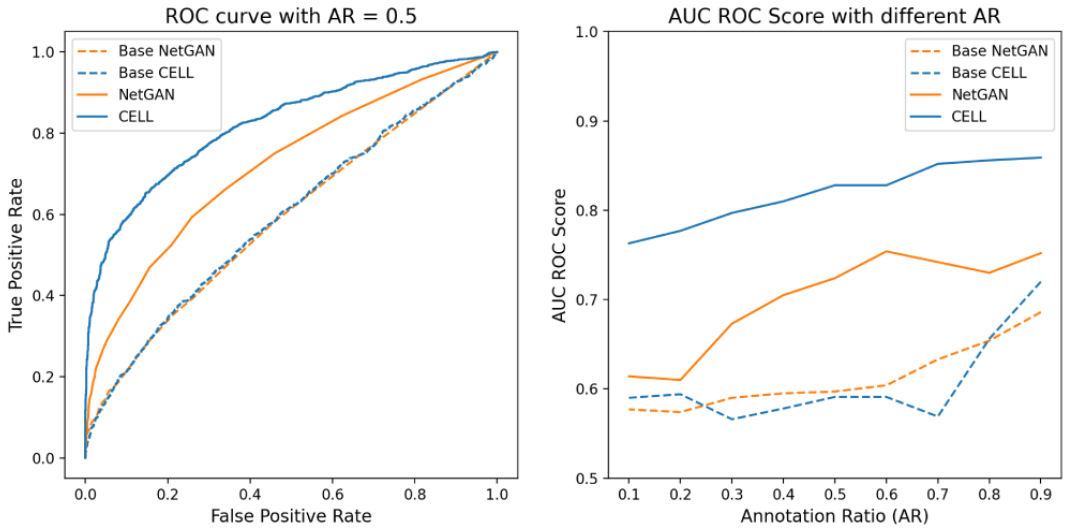
Results

Task Evaluation Outcomes

Task 1: Comparison of the Link Prediction Results in a Graph With/Without the Unlabeled Associations Among Different ARs

We conducted our experiments in two scenarios. One was the base case (dashed line in Figure 3) where we tested our models without using the unlabeled information, without explicitly stating it as the base case; all of our statements in the following section would be the default case (solid line in Figure 3) that indicated that we included the unlabeled associations in the link prediction tasks. We reported the AUC ROC score in Figure 3. Here, the left subfigure displayed the AUC ROC curve with a fixed AR of 0.5. The dashed line named “Base NetGAN” indicates the method of NetGAN that did not incorporate the unlabeled information. “Base CELL” is the method that CELL runs in the base case. There was little difference between the two methods when considering the base case with an AUC ROC score of 0.597 for NetGAN and 0.591 for CELL. However, when unlabeled information was taken into consideration, both methods achieved better performance compared to the base case (the AUC ROC score of NetGAN was 0.724, while CELL achieved a score of 0.828). The right side of Figure 3 shows the performance of the proposed methods in different ARs ranging from 0.1-0.9. We determined that CELL had overall better performance.

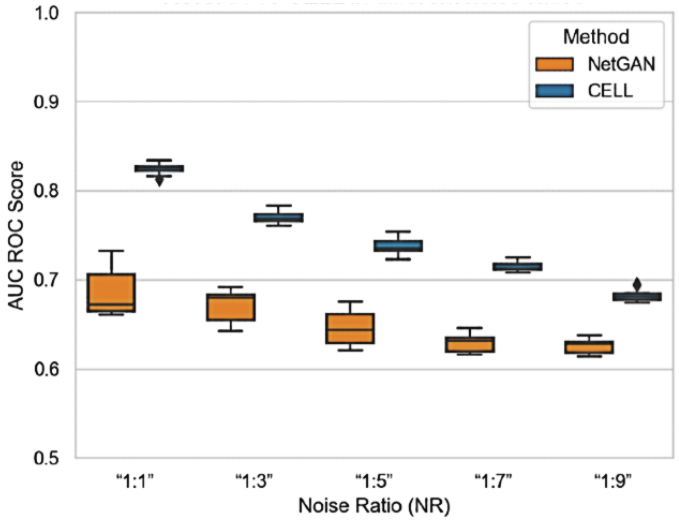
Figure 3. AUC ROC performance of NetGAN and CELL with/without unlabeled information. AR: annotation ratio; AUC ROC: area under the receiver operating characteristic curve; CELL: Cross-Entropy Low-rank Logits.



Task 2: How Do the Models Perform With Different NRs in the Unlabeled Edges?

We tested the performance of methods in which the unlabeled information contained a different ratio of noise 10 times (Figure 4). CELL demonstrated exceptional performance when the NR was 1:1. Even in the extreme case where the true versus false ratio reached 1:9, CELL still had better performance compared to NetGAN with an area under the curve (AUC) score of around 0.7. CELL had less variance in performance compared with NetGAN at all NRs. In other words, CELL had a relatively better capability and stability to use unlabeled data compared with NetGAN when dealing with the complexity of the NR in unknown information.

Figure 4. Performance in terms of AUC score at different noise ratios. AUC: area under the curve; AUC ROC: area under the receiver operating characteristic curve; CELL: Cross-Entropy Low-rank Logits.



Task 3: The Performance of Proposed Models in Our Collected Real Data Set

After our exploration of our methods in task 2, we conducted our methods on a real data set. Although the NR was unknown in our real data set, the proposed methods still performed better than random classification with the incorporation of unknown associations. In addition, compared with NetGAN, CELL still had an impressive result with an AUC ROC of up to 0.706 when the test and train ratio was 1:1 and the unknown association occupied about 99.95% as shown in Figure 5. The good performance of CELL showed that it had an excellent capability to predict the true association with the use of unlabeled data. We reported the AUC ROC value of each type of association separately in Figure 6. Combining Figure S2 in Multimedia Appendix 1 of edge degree of each type of association, we concluded that, as the degree is larger, there would be more noise contained in each edge. Thus, the results would be affected correspondingly. The average precision performance for our proposed models in our synthetic and real data sets can be found in Supplementary 3 in Multimedia Appendix 1.

Figure 5. Performance on real data set. ROC: receiver operating characteristic curve; CELL: Cross-Entropy Low-rank Logits.

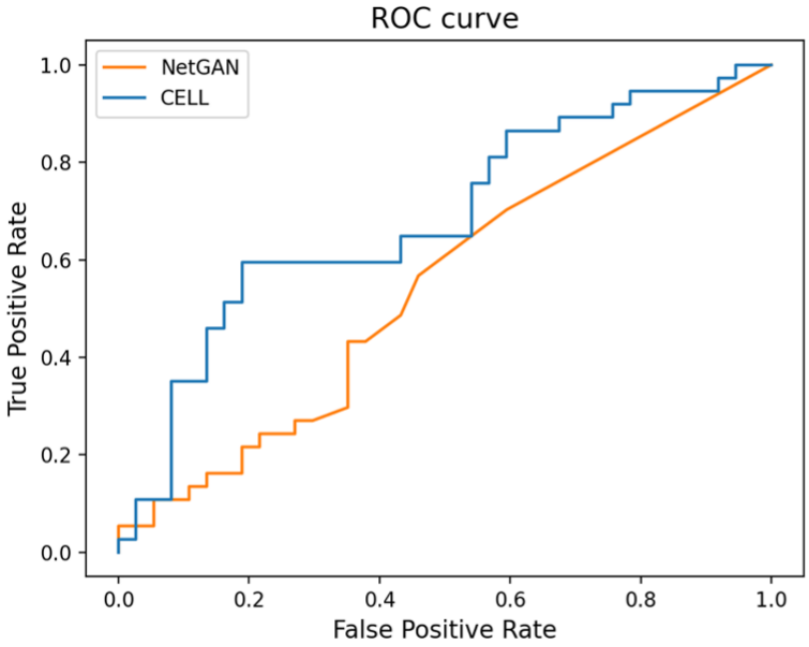
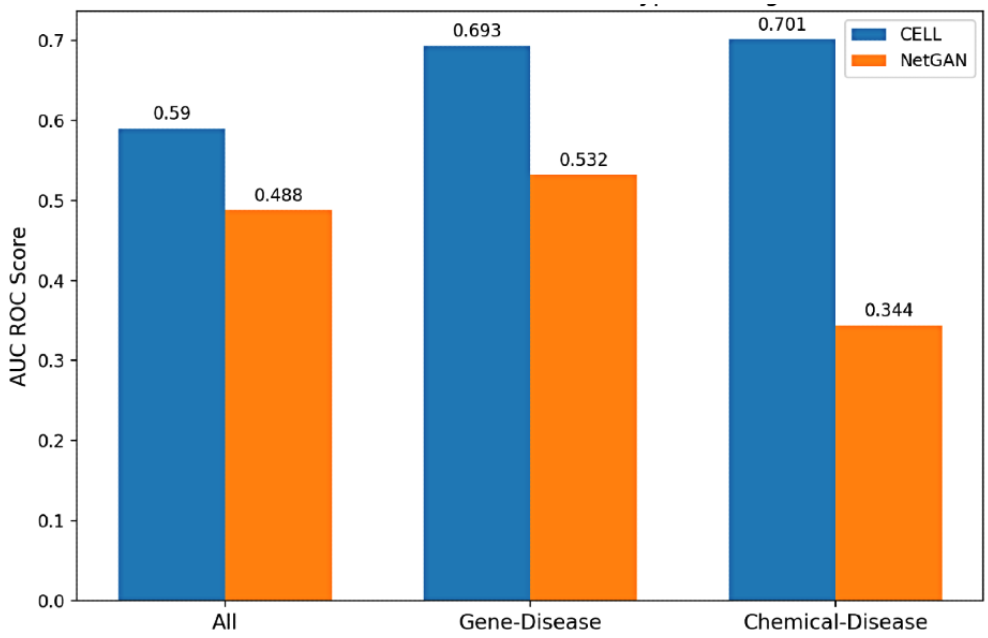


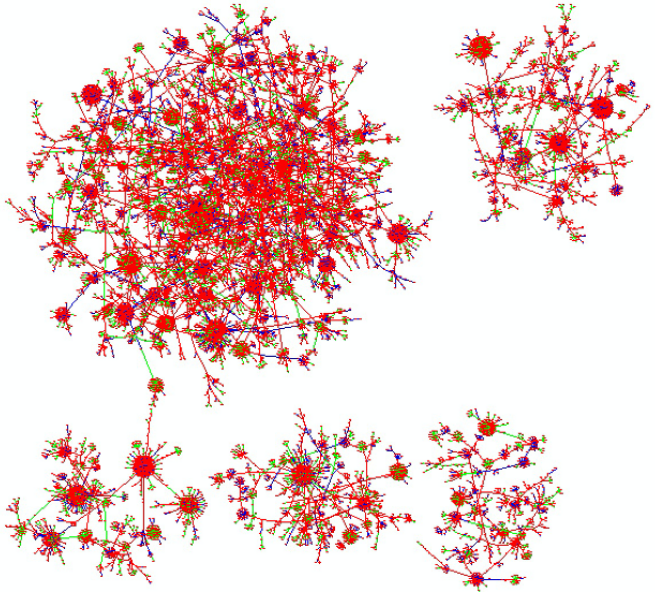
Figure 6. AUC ROC (area under the receiver operating characteristic curve) score for different types of associations in the real data set. CELL: Cross-Entropy Low-rank Logits.



Denoised Knowledge Graph Generated From the Real Data Set

We trained the adapted NetGAN with the whole real data set, and plotted the predicted denoised knowledge graph in Figure 7, where the edges are generated based on the score matrix calculated following the generation method used in NetGAN [12]. There are a total of 21,016 edges in our visualization consisting of gene-chemical (7562), gene-disease (7613), and chemical-disease (5841). Three different colors (red, green, and blue) stand for three different types of associations/edges (gene-chemical, gene-disease, chem-disease). The source file for the prediction can be found at [32].

Figure 7. Visualization of the predicted knowledge graph based on the real COVID-19–related data set.



Discussion

Principal Results

In this study, we proposed a method to automate the denoising of a knowledge graph generated via the counting of co-occurrence from biomedical literature. Our work can be considered as the preprocessing part for the curation of the knowledge graph. We adopted state-of-the-art generative-based graph methods, NetGAN and CELL, to leverage the unlabeled co-occurring biomedical entities in the training process by the perturbation of the original graph in the determination of an unknown edge. Two data sets (ie, synthetic and real data sets) were used to evaluate proposed methods in 3 link prediction tasks, and our experiments achieved promising results with both synthetic and real data sets.

Limitations

Despite the capability and stability of the methods used in this study, there are a few limitations that need to be discussed.

First, the associations labeled in the real data set are limited due to limited resources. In addition, due to the reality of vagueness or missing concepts in the biomedical literature, there will be some bias. A large sample of annotated associations may provide a solution to reduce this bias and thus is needed for our future work. One way to potentially accomplish this goal would be to use natural language processing methods to standardize the concepts prior to annotation, which may improve the construction of knowledge graph input to our methods. Another way includes collaborating with professional annotators to both increase the number of annotations as well as improve the quality.

Second, while we have achieved notable improvement with AUC around 0.7 in our real data set compared with random classification, there is still a gap between the experimental results in a controlled environment compared to the adaptation of the proposed methods for data curation in real-world scenarios. Performance improvement is still needed. The complexity of our investigated algorithm comes from the module of LSTM, which generates random walks for reconstructing the graph. An adaptation of binary neural networks [33] that directly produces the discrete adjacency matrix for the graph may have the potential to significantly improve the efficacy of our investigated methods as reconstruction of the adjacency matrix from random walks will not be needed. Another potential direction for improving the performance of removing noise in our investigated methods could be looking into the possibilities of transfer learning or external methods as we discussed previously, such as in [34]. By importing prior knowledge into the process of graph generation, we could employ the knowledge from an already built data set [35] to help us remove the false associations when constructing our biomedical graph.

Third, our evaluation was based on the logic of classifying the true or false associations directly, and was intentionally not focused on the impact evaluation of the denoised data sets generated in our work on downstream applications (eg, prediction for drug-target association and protein-protein interaction). Although we assume the performance will be improved in those applications [36], we acknowledge that there has not yet been any scientific proof to support that. The whole data stream, including the methods of data processing, data curation (ie, denoising method proposed in this study), and application, needs to be investigated further to fill this gap, which could provide convincing evidence of the impact of our proposed method for denoising knowledge base construction.

Acknowledgments

This work was supported by funding from the National Institutes of Health (NIH) National Institute of General Medical Sciences (NIGMS) (K99GM135488).

Authors' Contributions

NZ conceived and designed the study. CJ performed data integration, network construction, implementation of the algorithms, and experimentation, and created visualizations. NZ, CJ, and VN contributed to project implementation. NZ, CJ, and VN wrote the manuscript with contributions from all authors. All authors commented on and revised the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary materials.
[DOCX File , 411 KB - jmir_v24i7e38584_app1.docx]

References

1. Coronavirus disease (COVID-19). World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [accessed 2021-08-31]
2. Chen Q, Allot A, Lu Z. Keep up with the latest coronavirus research. Nature 2020 Mar 10;579(7798):193-193. [doi: 10.1038/d41586-020-00694-1] [Medline: 32157233]
3. El Mouden ZA, Taj RM, Jakimi A, Hajar M. Towards using graph analytics for tracking covid-19. Procedia Comput Sci 2020;177:204-211 [FREE Full text] [doi: 10.1016/j.procs.2020.10.029] [Medline: 33200008]
4. Groza A. Detecting fake news for the new coronavirus by reasoning on the Covid-19 ontology. ArXiv Preprint posted online on April 26, 2020 [FREE Full text]
5. CORD-19-on-FHIR. GitHub. URL: https://github.com/fhircat/CORD-19-on-FHIR [accessed 2021-09-27]
6. Oniani D, Jiang G, Liu H, Shen F. Constructing co-occurrence network embeddings to assist association extraction for COVID-19 and other coronavirus infectious diseases. J Am Med Inform Assoc 2020 Aug 01;27(8):1259-1267 [FREE Full text] [doi: 10.1093/jamia/ocaa117] [Medline: 32458963]
7. Zong N, Wong RSN, Yu Y, Wen A, Huang M, Li N. Drug-target prediction utilizing heterogeneous bio-linked network embeddings. Brief Bioinform 2021 Jan 18;22(1):568-580. [doi: 10.1093/bib/bbz147] [Medline: 31885036]
8. Zhou Y, Amimeur A, Jiang C, Dou D, Jin R, Wang P. Density-aware local Siamese autoencoder network embedding with autoencoder graph clustering. 2018 Oct 13 Presented at: 2018 IEEE International Conference on Big Data (Big Data); 2018; Seattle, WA, USA. [doi: 10.1109/bigdata.2018.8621992]
9. Zhou Y, Wu S, Jiang C, Zhang Z, Dou D, Jin R, et al. Density-adaptive local edge representation learning with generative adversarial network multi-label edge classification. 2018 Nov 17 Presented at: 2018 IEEE International Conference on Data Mining (ICDM); 2018; Singapore. [doi: 10.1109/icdm.2018.00203]
10. Yang T, Jiang C, Zhang Z, Dou D, Jin R, Wang P. Integrating local vertex/edge embedding via deep matrix fusion and siamese multi-label classification. 2019 Dec 9 Presented at: 2019 IEEE International Conference on Big Data (Big Data); 2019; Los Angeles, CA, USA. [doi: 10.1109/bigdata47090.2019.9006299]
11. Paulheim H. Knowledge graph refinement: A survey of approaches and evaluation methods. Semant Web 2016 Dec 06;8(3):489-508. [doi: 10.3233/sw-160218]
12. Pujara J, Miao H, Getoor L, Cohen W. Knowledge graph identification. 2013 Jan 1 Presented at: ISWC 2013; 2013; Sydney, Australia p. 542-557. [doi: 10.1007/978-3-642-41335-3_34]
13. Wienand D, Paulheim H. Detecting incorrect numerical data in DBpedia. 2014 Jan 1 Presented at: ESWC 2014; 2014; Paris, France p. 504-518. [doi: 10.1007/978-3-319-07443-6_34]
14. Long Y, Wu M, Liu Y, Fang Y, Kwoh CK, Chen J, et al. Pre-training graph neural networks for link prediction in biomedical networks. Bioinformatics 2022 Feb 16:2254-2262. [doi: 10.1093/bioinformatics/btac100] [Medline: 35171981]
15. Yang Q, Yan P, Zhang Y, Yu H, Shi Y, Mou X, et al. Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. IEEE Trans Med Imaging 2018 Jun;37(6):1348-1357. [doi: 10.1109/tmi.2018.2827462]
16. Zhou Z, Wang Y, Guo Y, Qi Y, Yu J. Image quality improvement of hand-held ultrasound devices with a two-stage generative adversarial network. IEEE Trans Biomed Eng 2020 Jan;67(1):298-311. [doi: 10.1109/tbme.2019.2912986]
17. Choi E, Biswal S, Malin B, Duke J, Stewart WF, Sun J. Generating multi-label discrete patient records using generative adversarial networks. 2017 Presented at: Machine Learning for Healthcare 2017; 2017; Boston, Massachusetts p. 305.
18. Li W, Wang Y, Cai Y, Arnold C, Zhao E, Yuan Y. Semi-supervised Rare Disease Detection Using Generative Adversarial Network. ArXiv Preprint posted online on December 3, 2018 [FREE Full text]

19. Rendsburg L, Heidrich H, Von Luxburg U. NetGAN without GAN: from random walks to low-rank approximations. 2020 Presented at: 37th International Conference on Machine Learning; 2020; Virtual.
20. Bojchevski A, Shchur O, Zügner D, Günnemann S. NetGAN: generating graphs via random walks. 2018 Presented at: 35th International Conference on Machine Learning; 2018; Stockholm, Sweden.
21. McCallum AK, Nigam K, Rennie J, Seymore K. Automating the construction of internet portals with machine learning. Information Retrieval 2000;3:127-163. [doi: 10.1023/A:1009953814988]
22. Wang Y, Yao Q, Kwok JT, Ni LM. Generalizing from a few examples. ACM Comput Surv 2021 May 31;53(3):1-34. [doi: 10.1145/3386252]
23. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997 Nov 15;9(8):1735-1780. [doi: 10.1162/neco.1997.9.8.1735] [Medline: 9377276]
24. Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks. 2017 Presented at: ICML'17: Proceedings of the 34th International Conference on Machine Learning; 2017; Sydney, NSW, Australia.
25. Grover A, Leskovec J. node2vec: Scalable feature learning for networks. 2016 Presented at: KDD '16; 2016; San Francisco, CA, USA. [doi: 10.1145/2939672.2939754]
26. Chen Q, Allot A, Lu Z. LitCovid: an open database of COVID-19 literature. Nucleic Acids Res 2021 Jan 08;49(D1):D1534-D1540 [FREE Full text] [doi: 10.1093/nar/gkaa952] [Medline: 33166392]
27. Wei CH, Kao HY, Lu Z. PubTator: a web-based text mining tool for assisting biocuration. Nucleic Acids Res 2013 Jul;41(Web Server issue):W518-W522 [FREE Full text] [doi: 10.1093/nar/gkt441] [Medline: 23703206]
28. Implementation of the paper 'NetGAN: Generating Graphs via Random Walks'. GitHub. URL: https://github.com/danielzuegner/netgan [accessed 2021-11-28]
29. Repository of the ICML 2020 paper NetGAN without GAN: From Random Walks to Low-Rank Approximations. GitHub. 2020. URL: https://github.com/hheidrich/CELL [accessed 2021-11-28]
30. scikit-learn: machine learning in Python. GitHub. URL: https://github.com/scikit-learn/scikit-learn [accessed 2021-11-28]
31. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003 Nov;13(11):2498-2504 [FREE Full text] [doi: 10.1101/gr.1239303] [Medline: 14597658]
32. Deep denoising of raw biomedical knowledge graph from COVID-19 literature, LitCovid and Pubtator. GitHub. URL: https://github.com/bioIKEA/deep_denoising_knowledge_graph [accessed 2021-10-04]
33. Dong HW, Yang YH. Training generative adversarial networks with binary neurons by end-to-end backpropagation. ArXiv Preprint posted online on October 10, 2018 [FREE Full text]
34. Jin W, Barzilay R, Jaakkola T. Junction tree variational autoencoder for molecular graph generation. 2018 Presented at: 35th International Conference on Machine Learning; 2018; Stockholm, Sweden.
35. Rossanez A, Dos Reis JC, da Silva Torres R, de Ribaupierre H. KGen: a knowledge graph generator from biomedical scientific literature. BMC Med Inform Decis Mak 2020 Dec 14;20(Suppl 4):314-328 [FREE Full text] [doi: 10.1186/s12911-020-01341-5] [Medline: 33317512]
36. Algan G, Ulusoy I. Image classification with deep learning in the presence of noisy labels: A survey. Knowledge-Based Systems 2021 Mar;215:106771. [doi: 10.1016/j.knosys.2021.106771]

Abbreviations

AP: average precision
AR: annotation ratio
AUC: area under the curve
AUC ROC: area under the receiver operating characteristic curve
CELL: Cross-Entropy Low-rank Logits
CORD-19: COVID-19 Open Research Dataset
GAN: generative adversarial network
LSTM: long short-term memory

Edited by C Basch; submitted 07.04.22; peer-reviewed by X Li, O Pavliuk; comments to author 01.05.22; revised version received 20.05.22; accepted 30.05.22; published 06.07.22.

Please cite as:
Jiang C, Ngo V, Chapman R, Yu Y, Liu H, Jiang G, Zong N
Deep Denoising of Raw Biomedical Knowledge Graph From COVID-19 Literature, LitCovid, and Pubtator: Framework Development and Validation
J Med Internet Res 2022;24(7):e38584
URL: https://www.jmir.org/2022/7/e38584
doi:10.2196/38584
PMID:35658098

Original Paper

Impact of COVID-19 Vaccine Misinformation on Social Media Virality: Content Analysis of Message Themes and Writing Strategies

Cindy Sing Bik Ngai[1], BA, MPhil, PhD; Rita Gill Singh[2], MSc, MEd, DSocSci; Le Yao[1], BA, MA

[1]Department of Chinese and Bilingual Studies, Hong Kong Polytechnic University, Kowloon, Hong Kong

[2]Language Centre, Hong Kong Baptist University, Kowloon, Hong Kong

Corresponding Author:
Cindy Sing Bik Ngai, BA, MPhil, PhD
Department of Chinese and Bilingual Studies
Hong Kong Polytechnic University
AG520b
Kowloon
Hong Kong
Phone: 852 27667465
Email: cindy.sb.ngai@polyu.edu.hk

Abstract

Background: Vaccines serve an integral role in containing pandemics, yet vaccine hesitancy is prevalent globally. One key reason for this hesitancy is the pervasiveness of misinformation on social media. Although considerable research attention has been drawn to how exposure to misinformation is closely associated with vaccine hesitancy, little scholarly attention has been given to the investigation or robust theorizing of the various content themes pertaining to antivaccine misinformation about COVID-19 and the writing strategies in which these content themes are manifested. Virality of such content on social media exhibited in the form of comments, shares, and reactions has practical implications for COVID-19 vaccine hesitancy.

Objective: We investigated whether there were differences in the content themes and writing strategies used to disseminate antivaccine misinformation about COVID-19 and their impact on virality on social media.

Methods: We constructed an antivaccine misinformation database from major social media platforms during September 2019-August 2021 to examine how misinformation exhibited in the form of content themes and how these themes manifested in writing were associated with virality in terms of likes, comments, and shares. Antivaccine misinformation was retrieved from two globally leading and widely cited fake news databases, COVID Global Misinformation Dashboard and International Fact-Checking Network Corona Virus Facts Alliance Database, which aim to track and debunk COVID-19 misinformation. We primarily focused on 140 Facebook posts, since most antivaccine misinformation posts on COVID-19 were found on Facebook. We then employed quantitative content analysis to examine the content themes (ie, safety concerns, conspiracy theories, efficacy concerns) and manifestation strategies of misinformation (ie, mimicking of news and scientific reports in terms of the format and language features, use of a conversational style, use of amplification) in these posts and their association with virality of misinformation in the form of likes, comments, and shares.

Results: Our study revealed that safety concern was the most prominent content theme and a negative predictor of likes and shares. Regarding the writing strategies manifested in content themes, a conversational style and mimicking of news and scientific reports via the format and language features were frequently employed in COVID-19 antivaccine misinformation, with the latter being a positive predictor of likes.

Conclusions: This study contributes to a richer research-informed understanding of which concerns about content theme and manifestation strategy need to be countered on antivaccine misinformation circulating on social media so that accurate information on COVID-19 vaccines can be disseminated to the public, ultimately reducing vaccine hesitancy. The liking of COVID-19 antivaccine posts that employ language features to mimic news or scientific reports is perturbing since a large audience can be reached on social media, potentially exacerbating the spread of misinformation and hampering global efforts to combat the virus.

(J Med Internet Res 2022;24(7):e37806) doi:10.2196/37806

KEYWORDS

antivaccine misinformation; content themes; writing strategies; COVID-19; virality; social media; content analysis

Introduction

Background

Although vaccines are safe and effective in preventing life-threatening diseases, vaccine hesitancy is still prevalent globally [1,2]. Vaccine hesitancy refers to a delay in acceptance or refusal to vaccinate despite the availability of vaccines [3]. Vaccine hesitancy hovers along the continuum between the two extreme poles of high vaccine demand and vaccine refusal [4]. Vaccine hesitancy is viewed as one of the top 10 threats to global health [5] as it can compromise the herd immunity required to contain pandemics and lead to a greater transmission of the virus [6,7], particularly hampering efforts to curtail the COVID-19 pandemic.

Many complex reasons relating to sociodemographic factors and public trust account for vaccine hesitancy [6,8,9], with misinformation being the main factor [1,7,10]. The World Health Organization (WHO) has used the term "infodemic" to refer to "the rapid spread of misleading or fabricated news" [11]. The topic of vaccination is subject to misinformation [12], particularly for newer vaccines [8], and the proliferation of misinformation, which has fueled fear about vaccine safety and its side effects, is regarded as the main cause of vaccine hesitancy [13,14]. Considerable research has found evidence of how misinformation about vaccines has led to lower vaccine intentions and uptakes [9,12,15-17].

It is worth explaining the distinctions between various terms that refer to misinformation. After the US presidential election in 2016, "fake news" as a phrase has gained considerable attention [18]. Fake news overlaps with other types of misleading information such as misinformation and disinformation. They can be distinguished primarily by the intent and mode of spread [18]. Misinformation is defined as "any health-related claim of fact that is (…) false or inaccurate due to a lack of scientific evidence" [19] and is shared by someone unwittingly without an intention to cause harm [20]. Misinformation specifically refers to claims that draw conclusions using incomplete or wrong information [21]. Conversely, disinformation refers to someone who deliberately creates and disseminates false information with an intention to cause harm [20]. While fake news has received substantial attention, it is difficult to define and has been used by some political groups to undermine certain news media [22]. Drawing on a previous study investigating health-related misinformation on social media [18], in this paper, we use the term *misinformation* as an umbrella term to refer to false or inaccurate health-related information about COVID-19 vaccines, irrespective of the intent, which is difficult to determine.

Impact of Misinformation on Social Media

Social media, recognized for its openness and participatory nature [23,24], is a common source to receive health information [25], share information about vaccines [16], and receive emotional support in crises [26]. Users can enhance their knowledge about a new disease, its transmission, and preventive measures [27]. However, at the same time, social media can be a source of widespread propagation of fake news, as users can post misinformed claims about vaccines, amplifying concerns about vaccines and resulting in increased vaccine hesitancy [12,15,16,28,29]. This poses a threat to public health and disrupts efforts to prevent disease via vaccines globally [5,29-31]. A recent study [32] highlighted that people exposed to vaccine information on social media have a higher proclivity to be misinformed and have vaccine hesitancy.

In the context of COVID-19, increased usage of social media is seen [33,34] alongside an increased amount of misinformation, negatively affecting public health [35]. Being exposed to COVID-19 information on social media has been linked to higher susceptibility to misinformation [36], resonating with the literature showing that the public is likely to be exposed to misinformation on social media [37]. A few factors have contributed to the rising influence of misinformation on COVID-19 vaccines on social media. These include the notion that lockdowns resulting from COVID-19 in many countries rendered people to have more time to access social media [34], thereby increasing the likelihood of exposure to misinformation. Moreover, since such news tends to be amusing and novel, it encouraged sharing behavior [38]. This has been evidenced in an observational study conducted in Italy, where 2000 articles posted on COVID-19 were analyzed, with articles containing misinformation shared 2 million times, constituting 78% of the total shares of all articles [39]. Another factor relates to many social media sites (eg, Twitter) adopting a strict limit on characters, meaning that the information presented may not be contextualized, making it misleading or incomplete [12]. In the early phase of the COVID-19 pandemic, social media companies did not adopt timely actions against misinformation on their sites [20].

Prior studies have documented evidence of misinformation about COVID-19 on social media [12,40,41]. Some typical examples are that more antivaccine messages were evident on Twitter than provaccine messages [12], while viewers were likely to encounter antivaccine videos on YouTube [40]. A poll conducted by Ofcom showed that 46% of British people reported having been exposed to misinformation about COVID-19, and of those who were exposed, approximately 66% reported watching these sources every day [42], thereby accelerating beliefs in misinformation because of repeated exposure [43]. Several studies have shown that the exposure to misinformation is closely associated with vaccine hesitancy (eg, [9,12,15]).

While there is a large body of research on the association between antivaccine misinformation and vaccine hesitancy [6,7,9,25,44,45], the specific content themes of discussion regarding antivaccine misinformation about COVID-19, how these content themes are manifested in writing through the use of certain writing strategies, and how these themes affect virality on social media warrant examination. Previous studies have so far mainly focused on content themes relating to misinformation

and their association with vaccine hesitancy via surveys and experimental studies without considering a range of writing strategies employed to disseminate these messages and the use of social media (eg, [6,7,9,45]). Recently, researchers have investigated these aspects on Facebook and Twitter [25,44]; however, insufficient attention has been paid to *how* content is manifested in writing via the use of certain strategies [12,46-48]. Studies often overlook the use of such strategies in antivaccine misinformation. Given that the public creates and extracts meanings from social media posts [49], the analysis of language, such as determining how content themes are manifested through the use of writing strategies, is critical in gaining a comprehensive understanding of the misinformation that is shared on social media.

Additionally, only a handful of studies have investigated the impact of misinformation on vaccine hesitancy as exhibited in virality in the form of comments, reactions, and shares on social media (eg, [12,17,46,50-53]). Specifically, one study focused on both pro- and antivaccine themes on Twitter from 2014 to 2017, noting that safety concerns and conspiracy theories were the most prevalent themes, and these themes were associated with sentiment-based opinions [12]. Another study analyzed negative and positive comments to Facebook posts on vaccine hesitancy related to human papillomavirus (HPV) in South Africa [17]. Other studies focused on HPV on Twitter and YouTube [50,51]. However, scarce attention has been accorded to the impact of misinformation related to COVID-19 vaccines on virality on social media. More importantly, the ways in which the content themes are manifested in the writing strategies that social media users employ [54] may facilitate the spread of misinformation [55], since the public comprehends the provided information based on their discursive resources [49]. This study is therefore deemed a worthwhile endeavor to undertake.

Developing an Integrated Framework of Antivaccine Misinformation on COVID-19

Strategy Overview

Based on the gaps identified above, in that insufficient attention has been paid to how content themes on antivaccine misinformation alongside how such themes are manifested in writing are associated with virality of misinformation on social media, particularly in the context of COVID-19, the need for this study was evident. Most studies have examined content themes on COVID-19 vaccine hesitancy and misinformation [25,44] without considering how these themes are manifested in writing through the use of writing strategies. Kata's [47] seminal work on a content analysis of antivaccination websites revealed not only the presence of a variety of content themes but also the writing strategies on these sites, namely amplification strategies, using credible sources untruthfully, misrepresenting facts and statistics, and personal testimonies. Her work has been drawn upon by scholars, which will be elaborated in the next section. A review of the literature (eg, [12,46,47]) indicates that certain writing strategies are used to make misinformation more credible. These strategies comprise: (1) mimicking the language features and format of mainstream news media and scientific reports, (2) using a conversational style such as a personal and informal tone of writing, and (3) employing amplification or exaggeration. Drawing on the above, we developed a framework by integrating two key dimensions, content themes on antivaccine misinformation and the writing strategies used to convey this information, to investigate how these aspects are associated with the virality of misinformation on social media. The findings can inform public health communication efforts with respect to how the public responds to these themes, and consequently offer targeted interventions from social media platform providers, health organizations, and governments to reduce COVID-19 vaccine hesitancy. In the following, we explain these two key dimensions of our study: content themes and writing strategies.

Content Themes

Theme Development

Following an extensive survey of the published literature, although vaccine hesitancy is influenced by many factors in different historical, political, and sociocultural contexts [8,56], the main attributional factors in terms of antivaccine misinformation content themes appear to be safety, conspiracy, and efficacy [12,14,25,47]. Therefore, we incorporated these content themes into our proposed framework.

Safety Concern

Studies show that vaccine safety concern is an important factor causing vaccine hesitancy [12,38,47,57]. Antivaccine safety concern is defined as content that discredits the safety of vaccines and may include notions that vaccines cause harm or death without providing immunity [12,47,58]. This concern is amplified by misinformation spread on social media, in particular that COVID-19 vaccines were developed very quickly, and are therefore unsafe and that all of the side effects have yet to be investigated [38]. In a recent study of parents in Australia, approximately 24% of participants were reluctant or not sure of getting a COVID-19 vaccine and of these, 89% had concerns about vaccine safety [59]. In high-income countries with effective vaccination programs, the fear of safety risks of vaccines far surpassed the fear of the diseases that vaccines prevent [60].

Conspiracy Theories

Conspiracy theories are associated with exposure to misinformation on social media [14,61]. This content theme presents specific conspiracy theories, which may encompass stories of fake claims of microchips and poison found in vaccines; fraud; collusion between pharmaceutical companies, governments, and doctors; and pharmaceutical companies manipulating data on vaccine efficacy to make huge profits [12,14,25,47]. A growing number of studies have shown that conspiracy beliefs are associated with vaccine hesitancy and uptake [6,7]. In the United States, beliefs in COVID-19 conspiracy theories were negatively related to the perceived safety of vaccinations and willingness to get vaccinated [45]. The explanation for this is the reduced perceptions of the threat and concerns about safety [62]. Conspiracy beliefs have a widespread influence and discourage vaccine uptake because they are difficult to counter, and are linked to a propensity to reject information from science experts [63]. Beliefs in one conspiracy theory are often tied to beliefs in others, indicating

that the public is more likely to trust these beliefs irrespective of their content [64].

Efficacy Concerns

This content theme presents vaccines as ineffective and unnecessary, emphasizing that they are unsuccessful and that an increased incidence in the disease is seen after vaccination [12,47]. For example, instead of preventing disease, it is believed that one is more susceptible to getting COVID-19 from the vaccine. In one instance, statistics were cited showing that most people contracting vaccine-preventable diseases (VPDs) were those who had been vaccinated, indicating that vaccination is ineffective [47]. In the study on parents in Australia cited above, approximately 24% of participants were reluctant to get a COVID-19 vaccine and of these, 89% had concerns about vaccine efficacy, believing that the vaccine was unnecessary [59].

The above studies on content themes generated the first research question (RQ1): Are there any differences in the content themes disseminated in COVID-19 antivaccine misinformation on social media?

Writing Strategies

Categorization

Kata's [47] pioneering study on content themes and writing strategies employed on antivaccine websites has been drawn on by several researchers. Given the prevalent use of social media in the last 10 years, such strategies might have undergone some changes. Jamison et al [12] built upon Kata's [47] typology in their analysis of the types of vaccine misinformation on Twitter from 2014 to 2017 by using a data set of 1.8 million tweets. They acknowledged that most of Kata's [47] content themes and writing strategies were still relevant, but they also observed that amplification strategies such as hashtags to promote content and @messages to high-profile people and organizations to gain attention were commonly seen [12]. Further, they found that antivaccine claims were frequently presented on Twitter as truths by mimicking the language of mainstream news or science [12]. The few studies documented on writing strategies used in antivaccine misinformation suggest that such misinformation is typified by a personal and conversational tone, as evidenced by the use of short texts to enhance comprehension by the public [12,48,65,66]. Personal experiences or anecdotes are emphasized to appeal to the public's emotions [34]. Drawing on these studies, we categorized writing strategies into the following three types.

Format and Features of Writing That Mimic the Language of Mainstream News or Science

Antivaccine claims can be posted as legitimate truths by imitating the language of mainstream news or science experts and presenting them in an accessible language to laypeople [12,46]. Previous research has shown that misinformation that is scientific-sounding is related to lower vaccine intentions [9]. In a Twitter study over the period of 2014-2017, some claims were found to be presented as facts by mimicking the language features used by science experts or the news [12]. In another study on 16,768 tweets on Twitter in 2018, statistics were distorted to support antivaccine claims [46]. In a study on antivaccine misinformation on websites in 2010, credible sources were used dishonestly, false conclusions were derived, and statistics were misrepresented [47]. For example, statistics were quoted showing that the majority of people who got VPDs had been vaccinated, demonstrating that vaccination was ineffective; however, statistics on the high number of unvaccinated people who contracted VPDs were not indicated [47]. Drawing on the findings of these studies, we argue that mimicking of mainstream news and scientific reports can be manifested through the use of writing strategies such as explaining actions taken by health institutions/medical experts; quoting from public figures; using jargon and statistics; attributing information to credible-sounding sources, including medical experts/health organizations and scientific studies; and capitalizing all letters of the first word in a sentence/heading.

Use of a Conversational Writing Style

Language can be utilized in different ways to express ideas. One way in which this is done is via the use of a conversational/personal tone of voice or a formal/impersonal tone [46,47]. The former notion is more relevant to antivaccine misinformation, as has been shown in studies where antivaccine misinformation is dominated by a conversational and personal tone as well as personal experiences/anecdotes, which induce fear, anxiety, and mistrust [12,48,65,66]. Personal experiences serve an important role in appealing to the public's emotions by instilling fear and using blame rather than appealing to logic [38,46]. Existing literature suggests that antivaccine messages often adopt a conversational style by using short sentences or texts, sentence fragments, and questions, facilitating the public's comprehension and making the language accessible to anybody [46,48,67]. Specifically, Italian webpages disseminating squalene-based influenza provaccine information had on average longer words and sentences that reduced their readability, whereas antivaccine webpages were easy to read [48]. Other researchers also found that in comparison to antiinfluenza immunization online messages, the proinfluenza immunization messages were more difficult to read due to their formal writing style [67]. Building on these studies, we argue that the conversational style and personal tone of voice is manifested in the use of informal expressions (eg, sentence fragments, questions, contractions, emojis), use of first- and second-person pronouns, author visibility, and sharing of personal experiences.

Use of Amplification

Amplification refers to how information can be distorted, amplified, or exaggerated on social media [46,68]. Antivaccine advocates have utilized Facebook and Twitter to disseminate exaggerated claims [12,46]. In a study on vaccine hesitancy on Twitter, it was found that most of the negative tweets on COVID-19 contained a hashtag as opposed to positive and neutral tweets [69]. Similarly, a study showed that antivaccine claims on Twitter in 2018 relied on the use of hashtags [46]. Another example pertains to the link between the measles-mumps-rubella vaccine and autism in the Wakefield study [70], which was retracted in 2010; however, Google Scholar statistics indicated that as of June 26, 2018, the Wakefield study had been cited 1090 times since 2012. It should be noted that some of these citations highlighted the flaws in the study, whereas other studies did not do so, suggesting

amplification [68]. Two amplification strategies are considered in this study. The first is the use of hashtags, which are popular on social media; in particular, content on Twitter tends to use a large number of antivaccine hashtags to amplify its messages [12,71]. The second frequently used amplification strategy considered is the use of @messages to celebrities and public figures to seek their attention [12].

Following this, the second research question (RQ2) posed is: Are there any differences in the writing strategies manifested in COVID-19 antivaccine misinformation on social media?

Virality of Social Media Posts

It is vital to examine the synergistic effect of antivaccine misinformation as exhibited in content and writing strategies on the virality of misinformation on social media. Virality is a term referring to the wide reach or attention of a social media activity or post [72,73]. Viral posts can reach a large audience [74], having far-reaching consequences. The literature has shown that virality can be observed from indicators such as likes, shares, favorites, and retweets on Twitter and Facebook [73,75]. Since our study focused on Facebook, we used the indicators likes, shares, and comments. Social media users use "likes" to indicate their interest in and attention to a topic [76], whereas a "share" is an indicator of user recommendation due to its extended communication [77]. A "comment" offers a platform for discussion since it requires the online user to reply to the post [78].

A rise in likes or shares for a post results in virality [73]. Some content themes attract substantial attention and become viral, increasing the likelihood that they will be shared with the public [79]. Previous studies have found mixed results on the type of content that is associated with virality. Positive and emotionally written articles that evoked strong emotions such as anger, and those with high practical value were more likely to be shared [80]. Yet, in another study, emotional posts had a negative relationship with virality on Twitter, Facebook, and Google, while posts with high practical utility were less often shared on Facebook [79]. Hansen et al [81] found that negative-news Tweets were more often retweeted. Additionally, antivaccine videos on HPV on YouTube led to more likes than provaccine videos [82,83].

Based on this previous work, the aim of this study was to examine the impact of antivaccine misinformation about COVID-19 on virality as exhibited in comments, shares, and likes.

The last research question (RQ) was thus derived as follows: What is the association between the content themes on COVID-19 antivaccine misinformation and the writing strategies used for the dissemination of this news on the virality of misinformation as exhibited in likes, comments, and shares?

Methods

Data Collection and Sample Period

We first constructed a database containing antivaccine misinformation circulating on social media for the examination of how COVID-19 misinformation exhibited in the form of content themes and manifested in writing strategies was associated with virality on social media. Antivaccine misinformation was retrieved from two prominent global fake news databases, International Fact-Checking Network (IFCN) Corona Virus Facts Alliance Database [84] and COVID Global Misinformation Dashboards [85], which aim to combat the infodemic by tracking and debunking COVID-19 misinformation [86]. The former was developed by the Corona Virus Facts Alliance, a committee under Poynter's IFCN, which covers COVID-19–related misinformation from fact-checkers in over 70 countries and in 43 languages of different text types funded by the Canadian Institutes of Health Research, Compute Canada, and the WHO. The latter is situated under the COVID-19 Misinformation Portal, which was developed and managed by the Social Media Lab at the Ted Rogers School of Management in Toronto. This portal tracks and visualizes coronavirus claims from more than 100 trusted fact-checkers. Both databases were developed by leading institutions and global organizations, having been widely cited in previous studies (eg, [87-90]), thus serving as reliable databases for sourcing data in this study.

To yield vaccine-related misinformation, we manually filtered vaccine-related misinformation using the keyword "vaccine" on the IFCN Corona Virus Facts Alliance Database and COVID Global Misinformation Dashboards from September 15, 2019, to August 16, 2021. In total, 2369 and 2298 fact-checked articles on "vaccine" were yielded from these two databases, respectively. Because these databases mainly publish review articles providing fact-checked reports on misinformation collected from multiple media sources (eg, online news, social media posts) in various languages, we trained a postgraduate student majoring in communication studies to carefully scrutinize 4667 vaccine review articles in these databases for retrieving the original links to the antivaccine misinformation on social media (eg, Facebook, Twitter, Instagram), although most links to the original sources were unavailable (ie, removed or deleted after being fact-checked).

To harvest the antivaccine misinformation on social media platforms that was available and comprehensible to social media users globally, two postgraduate students in communication studies were trained to manually visit and review 4667 fact-checked articles, as well as to check and retrieve the available original or archived posts in English and their related viral responses (likes, comments, and shares) on social media platforms. Finally, the trained students combined the yielded items from the two databases by removing the overlapping antivaccine fake news. YouTube was not included in the review process, since most original antivaccine misinformation videos had been removed.

In total, 350 posts containing misinformation on Facebook (n=285, 81.4%), Instagram (n=61, 17.4%), and Twitter (n=3, 0.8%) were yielded. As some posts only consisted of images and videos, and some were kept in an archive and thus some of their viral responses were unavailable, we filtered and retained posts that contained text messages (which included text-only posts and posts with image/video and text) and posts that generated virality in the form of likes, comments, and shares. Because most text-based posts were found in Facebook, the

most frequently used social media platform that has gained more active users in recent years [91,92], we decided to focus our text-based analysis on only Facebook posts in this study. Subsequently, we managed to capture 140 posts with all three indicators of virality (likes, comments, and shares) for further examination.

Content Analysis and Coding Scheme

Once our database had been constructed, we employed quantitative content analysis [34], a research method allowing researchers to conduct quantitative analysis on media messages in a scientific manner [93] to generate generalizable predictions [94] and draw conclusions [30,95]. Additionally, content analysis targets the context in which the occurrences of words, phrases, signs, and sentences are recorded, while offering an in-depth understanding [96]. As such, it is well-suited to a coding operation for a developed framework in media communication [30].

The coding scheme was developed based on the framework proposed in the previous section. Our framework consisted of two dimensions: the first dimension examined the content themes disseminated in misinformation posts and the second dimension focused on the writing strategies manifested in the content themes. The three subdimensions in the content themes included safety concerns (ie, posts that discredited the safety of vaccines), conspiracy theories (ie, posts that highlighted specific conspiracy theories), and efficacy concerns (ie, posts that advocated vaccines as ineffective and unnecessary). Three subdimensions were included in the writing strategies, namely mimicking the format or language features (ie, posts that mimicked the format and language features typical of real news or scientific reports), using a conversational style (ie, posts that were characterized by a conversational style and an informal, personal tone of voice), and using amplification (ie, posts that exaggerated the message by using hashtags and @messages to celebrities and public figures). Table 1 provides a description of the six subdimensions and their references.

Our examination of the data revealed that a post could contain multiple content themes to discredit vaccination. To minimize the loss of information, we coded the presence or absence of the subdimensions on a sentence basis [95]. For example, we coded the dominant subdimension in the content themes dimension to capture all content themes that were present when coding such posts. Textbox 1 shows a representative in-post text extracted from the database. This text first questions vaccine efficacy, suggesting that the vaccine is unnecessary and then continues to claim that the vaccine is unsafe due to its fatal side effects. Thus, the first and second sentence were coded as “efficacy concerns” and “safety concerns,” respectively.

Likewise, a post could employ more than one writing strategy. The first sentence in Textbox 2 mimicked a typical structure of fact-based news (eg, capitalizing all letters of the first word, describing actions of prominent staff from health institutions, using statistics), and was thus coded as format and language features that mimicked news media or scientific reports. The following three sentences adopted a different strategy, indicating a conversational style/personal tone of voice (eg, using sentence fragments; first-, second-, and third-person pronouns; contractions; and questions). Therefore, it was coded as a “conversational style.”

Since the post length varied from 1 to 17 sentences in the collected posts, we decided to normalize the data by dividing the number of sentences coded in each subdimension by the total number of sentences in each post.

Table 1. Description of the six subdimensions and their references.

Dimensions and subdimensions	Descriptions	References
Content themes		
Safety	Posts that discredit the safety of vaccines (eg, vaccines can cause harm or death)	[12,38,47,58,59]
Conspiracy	Posts that highlight specific conspiracy theories (eg, stories of fake claims of microchips found in vaccines; fraud; collusion between pharmaceutical companies, governments, and doctors; and pharmaceutical companies manipulating data to reap huge profits)	[6,12,14,25,45,47,61,63,64,97,98]
Efficacy	Posts that advocate vaccines as ineffective and unnecessary, emphasizing that they are unsuccessful and an increased incidence in the disease is seen after vaccination	[12,47,59]
Writing strategies		
Format and language features mimicking news or scientific reports	Posts that mimic the format and other features typical of real news or scientific reports. This is exhibited in the following ways: capitalizing all letters of the first word (eg, BREAKING, JUST IN); describing actions and quoting sentences from public figures; attributing information to credible-sounding sources, including medical experts, doctors/nurses, scientific studies, legal documents; using jargon, terminology, and/or statistics	[9,12,46,47,99]
Conversational style	Posts that are characterized by a conversational style or an informal, personal tone of voice. This is exhibited in: first- or second-person address form (eg, we should listen, you must act...); author visibility such as sharing personal experiences and feelings; and use of informal expressions (eg, using sentence fragments, questions, contractions, emojis, swear words)	[38,46,47]
Amplification	Strategies used to amplify or exaggerate the message. This is exhibited in the use of hashtags and @messages to celebrities and public figures	[12,46,68,70,71]

Textbox 1. Content themes present within one post.

- Are they really telling us that all 7,800,000,000 people in the world need to be vaccinated for a ‘virus’ that does not kill 99.99% of us?? …
- Reactions to the vaccine would kill more than the ‘virus’.

Textbox 2. Writing strategies present within one post.

- NEW: About 40-50% of CDC, FDA employees are refusing the COVID-19 vaccine according to Fauci, Marks — Breaking911 (@Breaking911) May 14, 2021.
- Double standards?
- What do they know they aren’t telling us?
- and You wonder why there’s no trust???????

Intercoder Reliability

Coding was performed by a doctoral student and a postgraduate student majoring in communication studies. Training was provided to both students by the first author before conducting the coding exercise, and the coders were invited to cocode 50 posts (ie, 30% of the total posts) during the training. The measure of intercoder reliability was calculated using the Cohen κ metric. The average Cohen κ of coded items was greater than 0.85, indicating almost perfect agreement [100].

Statistical Analyses

To fully reveal the weighting of specific content themes and writing strategies in each post, the counted number of sentences in each variable was divided by the total number of sentences in the corresponding post. We then employed analysis of variance (ANOVA) and the posthoc Tukey test to detect and compare the use of different content themes (RQ1) and writing strategies (RQ2) in antivaccine misinformation, since a previous study confirmed the robustness and validity of ANOVA in

testing the differences between independent variables, even if the normality assumption is violated [101].

In answering the last research question on the interaction between the content themes on antivaccine misinformation and the strategies used for the dissemination of this news on virality as exhibited in comments, reactions, and shares (RQ3), we first employed Poisson regression, a count regression model, in SPSS [102]. It was found that our data violated the assumption in Poisson regression due to an overdispersion of outcome variables, which is common in real-world data sets [103]. We therefore followed the common practice of replacing Poisson regression with negative binomial regression (NB2) to improve the goodness of fit, especially the Akaike information criterion and Bayesian information criterion [103]. NB2 is effective in fitting various types of data in communication and technical research, and is a more general model that relaxes the strong assumption that the underlying rate of the outcome is the same for each included participant [104].

Results

In response to RQ1, inquiring into whether there was any difference in the content themes disseminated in antivaccine misinformation on social media, the findings showed that safety concern was the most prominent theme, followed by conspiracy theories and efficacy (Table 2). The ANOVA results confirmed a significant difference in the content themes communicated in antivaccine misinformation ($F_{2,417}$=21.20, P<.001). The posthoc Tukey test indicated that the content theme safety concern was significantly higher than conspiracy theories (P=.003) and efficacy (P<.001), whereas conspiracy theories was also significantly higher than efficacy (P=.005). Table 2 provides the descriptive statistics on the examination of content themes disseminated in COVID-19 vaccine misinformation posts on Facebook and Figure 1 displays the mean count of sentences disseminating content themes in COVID-19 vaccine misinformation posts on Facebook.

Regarding RQ2, we investigated if there was any difference in the writing strategies employed to disseminate antivaccine misinformation. Our findings showed that conversational style was the most frequently used strategy, followed by format or language features mimicking news or scientific reports and amplification (Table 3). The ANOVA results revealed that there was a significant difference among the use of strategies ($F_{2,417}$=61.34, P<.001). The posthoc Tukey test confirmed that the conversational style strategy was significantly higher than format or language features mimicking news or scientific reports (P<.001) and amplification (P<.001), while format or language features mimicking news or scientific reports was also significantly higher than amplification (P<.001). See Table 3 for the descriptive statistics on the examination of writing strategies employed in COVID-19 vaccine misinformation posts on Facebook and Figure 2 for the mean count of sentences employing writing strategies in COVID-19 vaccine misinformation posts on Facebook.

Concerning RQ3, examining if there was any association between the content themes on antivaccine misinformation and the strategies used for the dissemination of this news on virality as exhibited in likes, comments, and shares, the NB2 results indicated that safety concern was a significant negative predictor of the number of likes and shares (Table 4). The odds ratio showed that for every extra sentence disseminating safety concerns, there was a decrease of 0.05 the number of likes and 0.30 the number of shares. By contrast, format or language features mimicking news or scientific reports was a strong positive predictor of the number of likes (Table 4). The odds ratio indicated that for every extra sentence utilizing format or language features mimicking news or science, there was an increase of 7.55 number of likes (Table 4).

The Omnibus test of NB2 showed significance with likes (P<.0001) and shares (P=.05) as a dependent variable, but not with comments as a dependent variable (P=.07).

Table 2. Descriptive statistics on the examination of content themes disseminated in COVID-19 vaccine misinformation posts on Facebook.

Content theme	Number of posts	Mean (SD)
Safety concern	140	0.23 (0.30)
Conspiracy theories	140	0.13 (0.25)
Vaccine efficacy	140	0.04 (0.14)
Total	420	0.13 (0.25)

Figure 1. Mean count of sentences disseminating content themes (CT) in COVID-19 vaccine misinformation posts on Facebook.

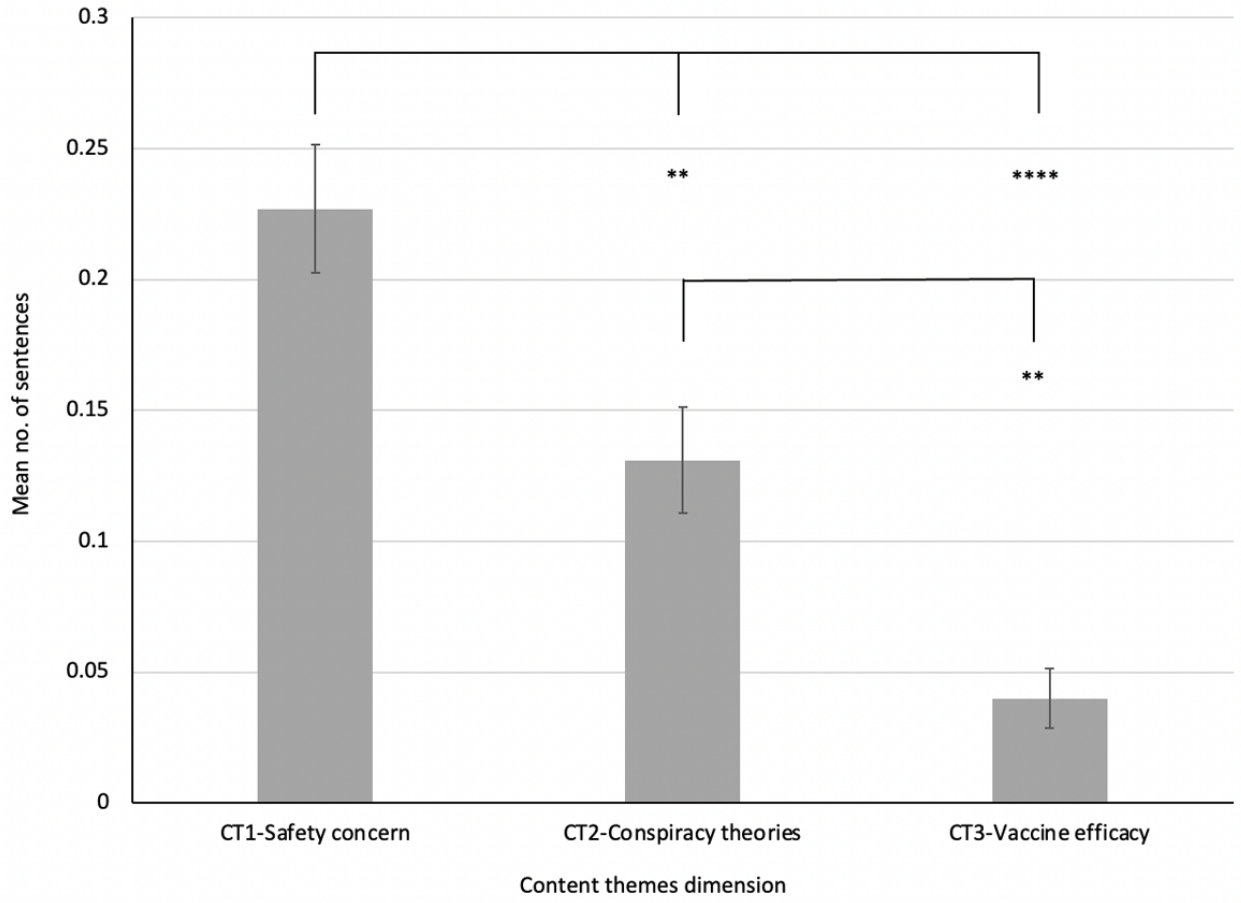


Table 3. Descriptive statistics on the examination of writing strategies employed in COVID-19 vaccine misinformation posts on Facebook.

Writing strategies	Number of posts	Mean (SD)
Format or language features mimicking news or scientific reports	140	0.29 (0.32)
Conversational style	140	0.45 (0.36)
Amplification	140	0.07 (0.16)
Total	420	0.27 (0.33)

Figure 2. Mean count of sentences employing writing strategies in COVID-19 vaccine misinformation (MS) posts on Facebook.

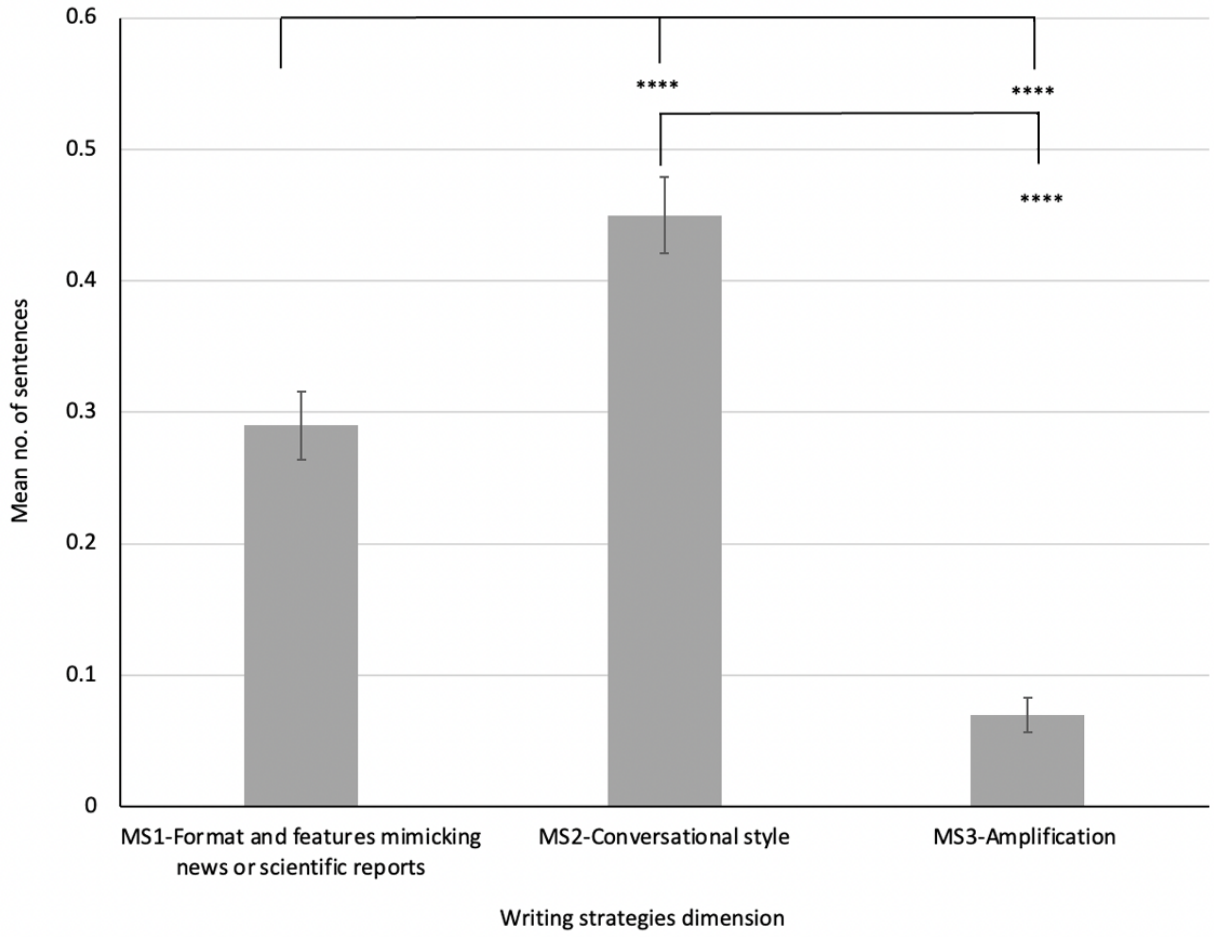


Table 4. Identification of positive and negative predictors of the numbers of likes, comments, and shares using a negative binomial regression model.

Variables	Likes			Comments			Shares		
	β (SE)	95% CI	*P* value	β (SE)	95% CI	*P* value	β (SE)	95% CI	*P* value
Content themes									
Safety concern	–3.04 (0.52)	–4.07 to –2.02	<.001	–1.52 (.48)	–2.45 to –.59	.001	–1.19 (.55)	–2.27 to –.12	.03
Conspiracy theories	.35 (.58)	–.79 to 1.49	.55	.06 (.51)	–.94 to 1.06	.90	.53 (.59)	–.63 to 1.70	.37
Efficacy	.22 (.83)	–1.40 to 1.84	.79	–.02 (.80)	–1.59 to 1.55	.98	–.32 (.89)	–2.06 to 1.42	.72
Writing strategies									
Format or language features mimicking news or scientific reports	2.02 (.93)	.19 to 3.85	.03	–.10 (.79)	–1.66 to 1.46	.90	–.20 (.75)	–1.68 to 1.28	.79
Conversational style	.23 (.74)	–1.21 to 1.68	.75	.23 (.69)	–1.12 to 1.58	.74	.26 (.73)	–1.18 to 1.70	.72
Amplification	.51 (1.16)	–1.76 to 2.77	.66	–.35 (1.03)	–2.36 to 1.67	.74	.70 (1.22)	–1.70 to 3.10	.57

Discussion

Principal Findings

The results showed that the most common content themes disseminated in COVID-19 antivaccine misinformation on Facebook were safety concerns, followed by conspiracy theories, which is consistent with previous studies [12,14,38,47]. A noteworthy point is the association between the content themes and virality of misinformation. Safety concern was a strong negative predictor of the number of likes and shares, although it was the most frequently used content theme. This could be attributed to the continued efforts made by governments and health organizations to emphasize the safety of COVID-19 vaccines and the growing threat of COVID-19 (eg, [105-107]). The public is therefore more likely to identify the misleading information disseminated in the posts, and less willing to like or share them when they have learnt more about the safety of vaccines. Earlier studies have shown that users are more likely to share content that has a high quality or practical value [80,108], which may also explain why the public was less willing to share this information.

With respect to the writing strategies manifested in the content themes, the results revealed that a conversational style as well as format and language features that mimicked news media and scientific reports were frequently used to spread antivaccine misinformation. This finding resonates with the literature about antivaccine information, which is typified by a personal, conversational, and negative tone [48,65,66], and a prior study showing that antivaccine claims were presented as facts by imitating the language of science and news on Twitter, leading to a high number of retweets [12]. Given that language can be utilized for different purposes [12,46,49,55], the antivaccine news posts capitalized on these distinct features of language to achieve their purpose of disseminating misinformation. Some prior studies have investigated strategies such as emotional appeal and amplification to disseminate antivaccine news [12,46,47]; however, studies on these aspects relating to COVID-19 on social media are lacking. Therefore, our findings add to the body of knowledge of how content themes are manifested in writing strategies to disseminate COVID-19 antivaccine misinformation.

Our results also confirmed that posts relying on format and language features that mimicked news media and scientific reports were strong positive predictors of likes. These posts might have looked authentic and appealing, thus encouraging liking. It is interesting that while the posts promoted liking, they were not associated with shares, possibly due to the negative information contained in them as well as uncertainty of the source of information, which might have made users hesitant to virally share the information. Facebook had a total of 2.91 billion monthly active users from October to December 2021 [91]; thus, the far-reaching effects of even liking antivaccine posts about COVID-19 should not be downplayed. These liked posts may exacerbate the extent of antivaccine misinformation disseminated on social media, potentially hampering efforts to prevent diseases via vaccines [5,32]. Our novel findings regarding the relationship between virality and the content themes and writing strategies used provide important insights for counteracting COVID-19 antivaccine misinformation.

Implications, Limitations, and Future Directions

This study contributes to the understanding of which content themes and writing strategies manifested in the themes led to virality of COVID-19 antivaccine misinformation on social media, and adds to the literature on this development subject. By constructing a database of antivaccine misinformation on COVID-19 circulating on social media from two globally leading and widely cited fake news databases for the examination of COVID-19 misinformation exhibited in the form of content themes and writing strategies and their association with virality, we found that posts on safety concerns were the most frequently occurring topic, and this content theme was manifested in writing through the use of a conversational style and format and language features that mimicked news and scientific reports. Additionally, the latter was associated with virality in the form of likes.

Our study thus provides insights into which content themes and manifestation strategies were associated with virality, and could be explored further to counter the impact of antivaccine misinformation. Since vaccine safety predicts vaccine intention, as found in other studies, and safety concern is the most frequently seen content theme susceptible to misinformation [109,110], the importance of countering misinformation on COVID-19 vaccines to increase public acceptance is confirmed. To do this successfully, systematic monitoring of the antivaccine misinformation circulating on social media has to be undertaken. This can be achieved by extracting misleading news posts related to safety and debunking the claims mentioned in these posts, especially those that adopt a conversational style and imitate real news or scientific reports. To discern real news and misinformation, social media platforms or fact-checkers should focus not only on the content but also how it is conveyed by paying more attention to the writing strategies used in such posts. It would be prudent for social media platform providers, governments, researchers, and health organizations to be provided with an updated summary of antivaccine misinformation circulating on social media to help them counter antivaccine concerns and provide accurate information about COVID-19 vaccines.

Like any data set, that used in this study has limitations. Since we only collected the antivaccine misinformation posts for 2 years, different time periods of the evolving COVID-19 pandemic should be considered. It should also be noted that antivaccine misinformation is subject to change over time, and thus our study findings should be interpreted with caution. The data on content themes and the writing strategies manifested in the content, and their associations with virality are correlational only. Most importantly, our study did not focus on social media users' sentiment-based opinions in the form of comments, which differ in valence (ie, negative, positive, neutral) and can reveal more detailed feelings [111]. An analysis of the valence of comments could have revealed more in-depth reasons that contributed to vaccine hesitancy in relation to antivaccine misinformation. Research suggests that emotions may overtake logic, and therefore studies have addressed antivaccine sentiment

impacts [13,112]. Our emphasis on COVID-19 content themes and writing strategies to disseminate such themes can be further empirically tested. Popular social media platforms have reached ubiquitous heights, and examination of other information-sharing platforms such as Instagram's COVID-19 antivaccine misinformation may shed more light on this topic.

Conclusions

To summarize, this study presents a novel examination of antivaccine misinformation in terms of content themes on COVID-19 and the ways in which these themes were manifested through the use of writing strategies. The key findings are that posts on safety concerns were negatively associated with likes and shares, whereas posts that mimicked the format and language features of news media and scientific reports were associated with likes on Facebook. This possibly suggests that antivaccine misinformation about COVID-19 has been amplified by liking these posts via social media. We do not yet know how far-reaching the impact of antivaccine misinformation has been, although some evidence indicates that misinformation about COVID-19 has had an impact on the public's vaccine uptake [113], posing a global health challenge. By drawing on this study's findings and leveraging the power of social media, platform providers, governments, and health organizations can take measures to counter COVID-19 antivaccine misinformation to reduce vaccine hesitancy, which remains pervasive globally.

Conflicts of Interest

None declared.

References

1. Carrieri V, Madio L, Principe F. Vaccine hesitancy and (fake) news: quasi-experimental evidence from Italy. Health Econ 2019 Nov;28(11):1377-1382 [FREE Full text] [doi: 10.1002/hec.3937] [Medline: 31429153]
2. Dubé E, Gagnon D, Nickels E, Jeram S, Schuster M. Mapping vaccine hesitancy--country-specific characteristics of a global phenomenon. Vaccine 2014 Nov 20;32(49):6649-6654 [FREE Full text] [doi: 10.1016/j.vaccine.2014.09.039] [Medline: 25280436]
3. The SAGE Working Group on Vaccine Hesitancy. Report of the Sage Working Group on vaccine hesitancy. Compass. 2014. URL: https://www.thecompassforsbc.org/sbcc-tools/report-sage-working-group-vaccine-hesitancy [accessed 2021-05-01]
4. 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]
5. Ten threats to global health in 2019. World Health Organization. 2019. URL: https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019 [accessed 2021-05-21]
6. Freeman D, Loe BS, Chadwick A, Vaccari C, Waite F, Rosebrock L, et al. COVID-19 vaccine hesitancy in the UK: the Oxford coronavirus explanations, attitudes, and narratives survey (Oceans) II. Psychol Med 2020 Dec 11:1-15 [FREE Full text] [doi: 10.1017/S0033291720005188] [Medline: 33305716]
7. Jolley D, Douglas KM. The effects of anti-vaccine conspiracy theories on vaccination intentions. PLoS One 2014;9(2):e89177 [FREE Full text] [doi: 10.1371/journal.pone.0089177] [Medline: 24586574]
8. Dubé E, Laberge C, Guay M, Bramadat P, Roy R, Bettinger J. Vaccine hesitancy: an overview. Hum Vaccin Immunother 2013 Aug;9(8):1763-1773 [FREE Full text] [doi: 10.4161/hv.24657] [Medline: 23584253]
9. Loomba S, de Figueiredo A, Piatek SJ, de Graaf K, Larson HJ. Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. Nat Hum Behav 2021 Mar;5(3):337-348. [doi: 10.1038/s41562-021-01056-1] [Medline: 33547453]
10. Verger P, Dubé E. Restoring confidence in vaccines in the COVID-19 era. Expert Rev Vaccines 2020 Nov;19(11):991-993. [doi: 10.1080/14760584.2020.1825945] [Medline: 32940574]
11. Immunizing the public against misinformation. World Health Organization. 2020. URL: https://www.who.int/news-room/feature-stories/detail/immunizing-the-public-against-misinformation [accessed 2021-11-01]
12. Jamison A, Broniatowski DA, Smith MC, Parikh KS, Malik A, Dredze M, et al. Adapting and extending a typology to identify vaccine misinformation on Twitter. Am J Public Health 2020 Oct;110(S3):S331-S339. [doi: 10.2105/AJPH.2020.305940] [Medline: 33001737]
13. 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]
14. Dubé E, Vivion M, MacDonald NE. Vaccine hesitancy, vaccine refusal and the anti-vaccine movement: influence, impact and implications. Expert Rev Vaccines 2015 Jan;14(1):99-117. [doi: 10.1586/14760584.2015.964212] [Medline: 25373435]
15. Basch CH, MacLean SA. A content analysis of HPV related posts on instagram. Hum Vaccin Immunother 2019;15(7-8):1476-1478 [FREE Full text] [doi: 10.1080/21645515.2018.1560774] [Medline: 30570379]
16. Ortiz RR, Smith A, Coyne-Beasley T. A systematic literature review to examine the potential for social media to impact HPV vaccine uptake and awareness, knowledge, and attitudes about HPV and HPV vaccination. Hum Vaccin Immunother 2019;15(7-8):1465-1475 [FREE Full text] [doi: 10.1080/21645515.2019.1581543] [Medline: 30779682]

17. Wiyeh AB, Cooper S, Jaca A, Mavundza E, Ndwandwe D, Wiysonge CS. Social media and HPV vaccination: unsolicited public comments on a Facebook post by the Western Cape Department of Health provide insights into determinants of vaccine hesitancy in South Africa. Vaccine 2019 Oct 08;37(43):6317-6323. [doi: 10.1016/j.vaccine.2019.09.019] [Medline: 31521412]
18. 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]
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. Wardle C, Singerman E. Too little, too late: social media companies' failure to tackle vaccine misinformation poses a real threat. BMJ 2021 Jan 21;372:n26. [doi: 10.1136/bmj.n26] [Medline: 33478950]
21. Southwell BG, Niederdeppe J, Cappella JN, Gaysynsky A, Kelley DE, Oh A, et al. Misinformation as a misunderstood challenge to public health. Am J Prev Med 2019 Aug;57(2):282-285. [doi: 10.1016/j.amepre.2019.03.009] [Medline: 31248741]
22. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science 2018 Mar 09;359(6380):1146-1151. [doi: 10.1126/science.aap9559] [Medline: 29590045]
23. Chen Q, Min C, Zhang W, Wang G, Ma X, Evans R. Unpacking the black box: How to promote citizen engagement through government social media during the COVID-19 crisis. Comput Human Behav 2020 Sep;110:106380 [FREE Full text] [doi: 10.1016/j.chb.2020.106380] [Medline: 32292239]
24. Kata A. Anti-vaccine activists, Web 2.0, and the postmodern paradigm--an overview of tactics and tropes used online by the anti-vaccination movement. Vaccine 2012 May 28;30(25):3778-3789. [doi: 10.1016/j.vaccine.2011.11.112] [Medline: 22172504]
25. Islam MS, Kamal AM, Kabir A, Southern DL, Khan SH, Hasan SMM, et al. COVID-19 vaccine rumors and conspiracy theories: the need for cognitive inoculation against misinformation to improve vaccine adherence. PLoS One 2021;16(5):e0251605 [FREE Full text] [doi: 10.1371/journal.pone.0251605] [Medline: 33979412]
26. Stephens KK, Malone PC. If the organizations won't give us information…: the use of multiple new media for crisis technical translation and dialogue. J Public Relat Res 2009 Apr 14;21(2):229-239. [doi: 10.1080/10627260802557605]
27. Lavorgna L, Ippolito D, Esposito S, Tedeschi G, Bonavita S. A disease in the age of the web: How to help people with Multiple Sclerosis in social media interaction. Mult Scler Relat Disord 2017 Oct;17:238-239. [doi: 10.1016/j.msard.2017.08.017] [Medline: 29055466]
28. Lahouati M, De Coucy A, Sarlangue J, Cazanave C. Spread of vaccine hesitancy in France: What about YouTube™? Vaccine 2020 Aug 10;38(36):5779-5782. [doi: 10.1016/j.vaccine.2020.07.002] [Medline: 32682617]
29. Zhang J, Featherstone JD, Calabrese C, Wojcieszak M. Effects of fact-checking social media vaccine misinformation on attitudes toward vaccines. Prev Med 2021 Apr;145:106408. [doi: 10.1016/j.ypmed.2020.106408] [Medline: 33388335]
30. Iacobucci G. Vaccination: "fake news" on social media may be harming UK uptake, report warns. BMJ 2019 Jan 23;364:l365. [doi: 10.1136/bmj.l365] [Medline: 30674463]
31. Larson HJ, Jarrett C, Eckersberger E, Smith DMD, Paterson P. Understanding vaccine hesitancy around vaccines and vaccination from a global perspective: a systematic review of published literature, 2007-2012. Vaccine 2014 Apr 17;32(19):2150-2159. [doi: 10.1016/j.vaccine.2014.01.081] [Medline: 24598724]
32. Stecula DA, Kuru O, Hall Jamieson K. How trust in experts and media use affect acceptance of common anti-vaccination claims. HKS Misinfo Review 2020 Jan 14;1(1):1-10. [doi: 10.37016/mr-2020-007]
33. Cortese A. Weibo reports record user growth in Q1, targets livestreaming and video content in 2020. KrASIA. 2020. URL: https://today.line.me/hk/v2/article/kNL8O1 [accessed 2020-05-22]
34. Ngai CSB, Singh RG, Lu W, Koon AC. Grappling with the COVID-19 health crisis: content analysis of communication strategies and their effects on public engagement on social media. J Med Internet Res 2020 Aug 24;22(8):e21360 [FREE Full text] [doi: 10.2196/21360] [Medline: 32750013]
35. Carrion-Alvarez D, Tijerina-Salina PX. Fake news in COVID-19: A perspective. Health Promot Perspect 2020 Nov 07;10(4):290-291 [FREE Full text] [doi: 10.34172/hpp.2020.44] [Medline: 33312921]
36. Johnson NF, Velásquez N, Restrepo NJ, Leahy R, Gabriel N, El Oud S, et al. The online competition between pro- and anti-vaccination views. Nature 2020 Jun;582(7811):230-233. [doi: 10.1038/s41586-020-2281-1] [Medline: 32499650]
37. Allen J, Howland B, Mobius M, Rothschild D, Watts DJ. Evaluating the fake news problem at the scale of the information ecosystem. Sci Adv 2020 Apr;6(14):eaay3539 [FREE Full text] [doi: 10.1126/sciadv.aay3539] [Medline: 32284969]
38. Lockyer B, Islam S, Rahman A, Dickerson J, Pickett K, Sheldon T, Bradford Institute for Health Research Covid-19 Scientific Advisory Group. Understanding COVID-19 misinformation and vaccine hesitancy in context: Findings from a qualitative study involving citizens in Bradford, UK. Health Expect 2021 Aug;24(4):1158-1167. [doi: 10.1111/hex.13240] [Medline: 33942948]
39. Moscadelli A, Albora G, Biamonte MA, Giorgetti D, Innocenzio M, Paoli S, et al. Fake news and Covid-19 in Italy: results of a quantitative observational study. Int J Environ Res Public Health 2020 Aug 12;17(16):5850 [FREE Full text] [doi: 10.3390/ijerph17165850] [Medline: 32806772]

40. Tang L, Fujimoto K, Amith MT, Cunningham R, Costantini RA, York F, et al. "Down the rabbit hole" of vaccine misinformation on YouTube: network exposure study. J Med Internet Res 2021 Jan 05;23(1):e23262 [FREE Full text] [doi: 10.2196/23262] [Medline: 33399543]
41. Wu W, Lyu H, Luo J. Characterizing discourse about COVID-19 vaccines: a Reddit version of the pandemic story. Health Data Science 2021 Aug 27;2021:9837856. [doi: 10.34133/2021/9837856]
42. Ofcom. Half of UK adults exposed to false claims about coronavirus. Ofcom. 2020 Apr 09. URL: https://www.ofcom.org.uk/about-ofcom/latest/features-and-news/half-of-uk-adults-exposed-to-false-claims-about-coronavirus [accessed 2021-05-01]
43. Pennycook G, Cannon TD, Rand DG. Prior exposure increases perceived accuracy of fake news. J Exp Psychol Gen 2018 Dec;147(12):1865-1880 [FREE Full text] [doi: 10.1037/xge0000465] [Medline: 30247057]
44. Lyu H, Zheng Z, Luo J. Misinformation versus facts: understanding the influence of news regarding COVID-19 vaccines on vaccine uptake. arXiv. 2022 Feb. URL: https://arxiv.org/abs/2106.07435 [accessed 2022-06-27]
45. Romer D, Jamieson KH. Conspiracy theories as barriers to controlling the spread of COVID-19 in the U.S. Soc Sci Med 2020 Oct;263:113356 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113356] [Medline: 32967786]
46. Maci S. Discourse strategies of fake news in the anti-vax campaign. LCM J 2019;6(1):14-43. [doi: 10.7358/lcm-2019-001-maci]
47. Kata A. A postmodern Pandora's box: anti-vaccination misinformation on the internet. Vaccine 2010 Feb 17;28(7):1709-1716. [doi: 10.1016/j.vaccine.2009.12.022] [Medline: 20045099]
48. Panatto D, Amicizia D, Arata L, Lai PL, Gasparini R. A comprehensive analysis of Italian web pages mentioning squalene-based influenza vaccine adjuvants reveals a high prevalence of misinformation. Hum Vaccin Immunother 2018 Apr 03;14(4):969-977 [FREE Full text] [doi: 10.1080/21645515.2017.1407483] [Medline: 29172967]
49. Marwick AE. Why do people share fake news? A sociotechnical model of media effects. Geor L Tech Rev 2018;474(2):512.
50. 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]
51. Kang GJ, Ewing-Nelson SR, Mackey L, Schlitt JT, Marathe A, Abbas KM, et al. Semantic network analysis of vaccine sentiment in online social media. Vaccine 2017 Jun 22;35(29):3621-3638 [FREE Full text] [doi: 10.1016/j.vaccine.2017.05.052] [Medline: 28554500]
52. Massey PM, Leader A, Yom-Tov E, Budenz A, Fisher K, Klassen AC. Applying multiple data collection tools to quantify human papillomavirus vaccine communication on Twitter. J Med Internet Res 2016 Dec 05;18(12):e318 [FREE Full text] [doi: 10.2196/jmir.6670] [Medline: 27919863]
53. Villar-Rodríguez G, Souto-Rico M, Martín A. Virality, only the tip of the iceberg: ways of spread and interaction around COVID-19 misinformation in Twitter. Commun Soc 2022 Apr 01;35(2):239-256. [doi: 10.15581/003.35.2.239-256]
54. Hewett K, Rand W, Rust RT, van Heerde HJ. Brand buzz in the echoverse. J Market 2016 May 01;80(3):1-24. [doi: 10.1509/jm.15.0033]
55. Di Domenico G, Tuan A, Visentin M. Linguistic drivers of misinformation diffusion on social media during the COVID-19 pandemic. Ital J Mark 2021 May 29;2021(4):351-369. [doi: 10.1007/s43039-021-00026-9]
56. Geoghegan S, O'Callaghan KP, Offit PA. Vaccine safety: myths and misinformation. Front Microbiol 2020;11:372. [doi: 10.3389/fmicb.2020.00372] [Medline: 32256465]
57. Larson HJ, Smith DMD, Paterson P, Cumming M, Eckersberger E, Freifeld CC, et al. Measuring vaccine confidence: analysis of data obtained by a media surveillance system used to analyse public concerns about vaccines. Lancet Infect Dis 2013 Jul;13(7):606-613. [doi: 10.1016/S1473-3099(13)70108-7] [Medline: 23676442]
58. Teoh D. The power of social media for HPV vaccination-not fake news!. Am Soc Clin Oncol Educ Book 2019 Jan;39:75-78 [FREE Full text] [doi: 10.1200/EDBK_239363] [Medline: 31099637]
59. Rhodes A, Hoq M, Measey M, Danchin M. Intention to vaccinate against COVID-19 in Australia. Lancet Infect Dis 2021 May;21(5):e110 [FREE Full text] [doi: 10.1016/S1473-3099(20)30724-6] [Medline: 32941786]
60. Larson HJ, Cooper LZ, Eskola J, Katz SL, Ratzan S. Addressing the vaccine confidence gap. Lancet 2011 Aug 06;378(9790):526-535. [doi: 10.1016/S0140-6736(11)60678-8] [Medline: 21664679]
61. Featherstone JD, Bell RA, Ruiz JB. Relationship of people's sources of health information and political ideology with acceptance of conspiratorial beliefs about vaccines. Vaccine 2019 May 21;37(23):2993-2997. [doi: 10.1016/j.vaccine.2019.04.063] [Medline: 31031028]
62. Jolley D, Douglas KM. Prevention is better than cure: Addressing anti-vaccine conspiracy theories. J Appl Soc Psychol 2017 Jun 28;47(8):459-469. [doi: 10.1111/jasp.12453]
63. Uscinski JE, Klofstad C, Atkinson MD. What drives conspiratorial beliefs? The role of informational cues and predispositions. Polit Res Quart 2016 Jan 13;69(1):57-71. [doi: 10.1177/1065912915621621]
64. Douglas KM, Uscinski JE, Sutton RM, Cichocka A, Nefes T, Ang CS, et al. Understanding conspiracy theories. Polit Psychol 2019 Mar 20;40(S1):3-35. [doi: 10.1111/pops.12568]
65. Bessi A, Zollo F, Del Vicario M, Scala A, Caldarelli G, Quattrociocchi W. Trend of narratives in the age of misinformation. PLoS One 2015;10(8):e0134641 [FREE Full text] [doi: 10.1371/journal.pone.0134641] [Medline: 26275043]

66. Hughes B, Miller-Idriss C, Piltch-Loeb R, Goldberg B, White K, Criezis M, et al. Development of a codebook of online anti-vaccination rhetoric to manage COVID-19 vaccine misinformation. Int J Environ Res Public Health 2021 Jul 15;18(14):7556 [FREE Full text] [doi: 10.3390/ijerph18147556] [Medline: 34300005]
67. Okuhara T, Ishikawa H, Okada M, Kato M, Kiuchi T. A readability comparison of anti- versus pro-influenza vaccination online messages in Japan. Prev Med Rep 2017 Jun;6:47-52 [FREE Full text] [doi: 10.1016/j.pmedr.2017.02.013] [Medline: 28271020]
68. Peters A, Tartari E, Lotfinejad N, Parneix P, Pittet D. Fighting the good fight: the fallout of fake news in infection prevention and why context matters. J Hosp Infect 2018 Dec;100(4):365-370. [doi: 10.1016/j.jhin.2018.08.001] [Medline: 30092290]
69. Roe C, Lowe M, Williams B, Miller C. Public perception of SARS-CoV-2 vaccinations on social media: questionnaire and sentiment analysis. Int J Environ Res Public Health 2021 Dec 10;18(24):13028 [FREE Full text] [doi: 10.3390/ijerph182413028] [Medline: 34948638]
70. Wakefield A, Murch S, Anthony A, Linnell J, Casson D, Malik M, et al. RETRACTED: 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]
71. Gargiulo F, Cafiero F, Guille-Escuret P, Seror V, Ward JK. Asymmetric participation of defenders and critics of vaccines to debates on French-speaking Twitter. Sci Rep 2020 Apr 20;10(1):6599. [doi: 10.1038/s41598-020-62880-5] [Medline: 32313016]
72. Han Y, Lappas T, Sabnis G. The importance of interactions between content characteristics and creator characteristics for studying virality in social media. Inf Syst Res 2020 Jun;31(2):576-588. [doi: 10.1287/isre.2019.0903]
73. Joseph N, Kar AK, Ilavarasan PV. How do network attributes impact information virality in social networks? Inf Disc Deliv 2021 Mar 22;49(2):162-173. [doi: 10.1108/idd-08-2020-0094]
74. Saha K, Dhar P. Reaping the benefits of virality in advertising - a Study of the Indian scenario. Asia Pac J Manag Technol 2020 Jul 1;1(1):24-29. [doi: 10.46977/apjmt.2020.v01i01.005]
75. Grover P, Kar AK, Davies G. “Technology enabled Health” – Insights from Twitter analytics with a socio-technical perspective. Int J Inf Manag 2018 Dec;43:85-97. [doi: 10.1016/j.ijinfomgt.2018.07.003]
76. Gittelman S, Lange V, Gotway Crawford CA, Okoro CA, Lieb E, Dhingra SS, et al. A new source of data for public health surveillance: Facebook likes. J Med Internet Res 2015 Apr 20;17(4):e98 [FREE Full text] [doi: 10.2196/jmir.3970] [Medline: 25895907]
77. Men LR, Tsai WS, Chen ZF, Ji YG. Social presence and digital dialogic communication: engagement lessons from top social CEOs. J Public Relat Res 2018 Jul 30;30(3):83-99. [doi: 10.1080/1062726x.2018.1498341]
78. Cho M, Schweickart T, Haase A. Public engagement with nonprofit organizations on Facebook. Public Relat Rev 2014 Sep;40(3):565-567. [doi: 10.1016/j.pubrev.2014.01.008]
79. Heimbach I, Schiller B, Strufe T, Hinz O. Content virality on online social networksmpirical evidence from Twitter, Facebook, and Google+ on German news websites. 2015 Presented at: 26th ACM Conference on Hypertext & Social Media; September 2015; Guzelyurt, Northern Cyprus p. 39-47. [doi: 10.1145/2700171.2791032]
80. Berger J, Milkman KL. What makes online content viral? J Market Res 2012 Apr 01;49(2):192-205. [doi: 10.1509/jmr.10.0353]
81. Hansen L, Arvidsson A, Nielsen F, Colleoni E, Etter M. Good friends, bad news-affect and virality in Twitter. 2011 Presented at: 6th International Conference on Future Information Technology; June 28-30, 2011; Crete, Greece URL: https://arxiv.org/pdf/1101.0510.pdf&embedded=true [doi: 10.1007/978-3-642-22309-9_5]
82. Briones R, Nan X, Madden K, Waks L. When vaccines go viral: an analysis of HPV vaccine coverage on YouTube. Health Commun 2012;27(5):478-485. [doi: 10.1080/10410236.2011.610258] [Medline: 22029723]
83. Yiannakoulias N, Slavik CE, Chase M. Expressions of pro- and anti-vaccine sentiment on YouTube. Vaccine 2019 Apr 03;37(15):2057-2064. [doi: 10.1016/j.vaccine.2019.03.001] [Medline: 30862365]
84. Fighting the infodemic: The #CoronaVirusFacts Alliance. Poynter. URL: https://www.poynter.org/coronavirusfactsalliance/ [accessed 2021-12-10]
85. COVID Global Misinformation Dashboards. URL: https://covid19misinfo.org/misinfowatch/ [accessed 2021-12-10]
86. COVID19MisInfo. URL: https://covid19misinfo.org/ [accessed 2022-05-30]
87. Coombes R, Davies M. Facebook versus the BMJ: when fact checking goes wrong. BMJ 2022 Jan 19;376:o95. [doi: 10.1136/bmj.o95] [Medline: 35045953]
88. Gruzd A, Mai P. The state of Social Media in Canada 2020. Social Media Lab. 2020 Jul 13. URL: https://socialmedialab.ca/2020/07/13/the-state-of-social-media-in-canada-2020-a-new-survey-report-from-the-ryerson-social-media-lab/ [accessed 2022-05-30]
89. Luengo M, García-Marín D. The performance of truth: politicians, fact-checking journalism, and the struggle to tackle COVID-19 misinformation. Am J Cult Sociol 2020 Sep 28;8:405-427 [FREE Full text] [doi: 10.1057/s41290-020-00115-w] [Medline: 33014362]
90. Naeem SB, Bhatti R. The Covid-19 'infodemic': a new front for information professionals. Health Info Libr J 2020 Sep;37(3):233-239 [FREE Full text] [doi: 10.1111/hir.12311] [Medline: 32533803]

91. Number of monthly active Facebook users worldwide as of 1st quarter 2022. Statista. URL: https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/#:~:text=How%20many%20users%20does%20Facebook,used%20online%20social%20network%20worldwide [accessed 2022-02-01]
92. Williamson DA. Global Facebook users 2020. eMarketer. 2020 Dec 08. URL: https://www.emarketer.com/content/global-facebook-users-2020 [accessed 2022-05-30]
93. Bock A, Isermann H, Knieper T. Quantitative content analysis of the visual. In: Margolis E, Pauwels L, editors. The Sage handbook of visual research methods. Newbury Park, CA: SAGE Publishing; 2011:265.
94. Krippendorff K. Content analysis: an introduction to its methodology, 2nd edition. Newbury Park, CA: Sage Publishing; 2003.
95. Boettger RK, Palmer LA. Quantitative content analysis: its use in technical communication. IEEE Trans Profess Commun 2010 Dec;53(4):346-357. [doi: 10.1109/tpc.2010.2077450]
96. Downe-Wamboldt B. Content analysis: method, applications, and issues. Health Care Women Int 1992;13(3):313-321. [doi: 10.1080/07399339209516006] [Medline: 1399871]
97. Hornsey MJ, Finlayson M, Chatwood G, Begeny CT. Donald Trump and vaccination: The effect of political identity, conspiracist ideation and presidential tweets on vaccine hesitancy. J Exp Soc Psychol 2020 May;88:103947. [doi: 10.1016/j.jesp.2019.103947]
98. Tanne JH, Hayasaki E, Zastrow M, Pulla P, Smith P, Rada AG. Covid-19: how doctors and healthcare systems are tackling coronavirus worldwide. BMJ 2020 Mar 18;368:m1090. [doi: 10.1136/bmj.m1090] [Medline: 32188598]
99. Shelby A, Ernst K. Story and science: how providers and parents can utilize storytelling to combat anti-vaccine misinformation. Hum Vaccin Immunother 2013 Aug;9(8):1795-1801 [FREE Full text] [doi: 10.4161/hv.24828] [Medline: 23811786]
100. Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. Tutor Quant Methods Psychol 2012;8(1):23-34 [FREE Full text] [doi: 10.20982/tqmp.08.1.p023] [Medline: 22833776]
101. Blanca M, Alarcón R, Arnau J, Bono R, Bendayan R. Non-normal data: Is ANOVA still a valid option? Psicothema 2017 Nov;29(4):552-557. [doi: 10.7334/psicothema2016.383] [Medline: 29048317]
102. Herbison P, Robertson MC, McKenzie JE. Do alternative methods for analysing count data produce similar estimates? Implications for meta-analyses. Syst Rev 2015 Nov 17;4:163 [FREE Full text] [doi: 10.1186/s13643-015-0144-x] [Medline: 26577545]
103. Sroka CJ, Nagaraja HN. Odds ratios from logistic, geometric, Poisson, and negative binomial regression models. BMC Med Res Methodol 2018 Oct 20;18(1):112 [FREE Full text] [doi: 10.1186/s12874-018-0568-9] [Medline: 30342488]
104. Gurland J. Some applications of the negative binomial and other contagious distributions. Am J Public Health Nations Health 1959 Oct;49:1388-1399. [doi: 10.2105/ajph.49.10.1388] [Medline: 13855879]
105. Edwards B, Biddle N, Gray M, Sollis K. COVID-19 vaccine hesitancy and resistance: Correlates in a nationally representative longitudinal survey of the Australian population. PLoS One 2021;16(3):e0248892 [FREE Full text] [doi: 10.1371/journal.pone.0248892] [Medline: 33760836]
106. Elgendy MO, Abdelrahim MEA. Public awareness about coronavirus vaccine, vaccine acceptance, and hesitancy. J Med Virol 2021 Dec;93(12):6535-6543 [FREE Full text] [doi: 10.1002/jmv.27199] [Medline: 34255346]
107. Schoch-Spana M, Brunson EK, Long R, Ruth A, Ravi SJ, Trotochaud M, et al. The public's role in COVID-19 vaccination: Human-centered recommendations to enhance pandemic vaccine awareness, access, and acceptance in the United States. Vaccine 2021 Sep 24;39(40):6004-6012 [FREE Full text] [doi: 10.1016/j.vaccine.2020.10.059] [Medline: 33160755]
108. Koohikamali M, Sidorova A. Information re-sharing on social network sites in the age of fake news. InformingSciJ 2017;20:215-235. [doi: 10.28945/3871]
109. Tavolacci MP, Dechelotte P, Ladner J. COVID-19 vaccine acceptance, hesitancy, and resistancy among university students in France. Vaccines 2021 Jun 15;9(6):654 [FREE Full text] [doi: 10.3390/vaccines9060654] [Medline: 34203847]
110. Wang C, Han B, Zhao T, Liu H, Liu B, Chen L, et al. Vaccination willingness, vaccine hesitancy, and estimated coverage at the first round of COVID-19 vaccination in China: A national cross-sectional study. Vaccine 2021 May 18;39(21):2833-2842 [FREE Full text] [doi: 10.1016/j.vaccine.2021.04.020] [Medline: 33896661]
111. Ji YG, Li C, North M, Liu J. Staking reputation on stakeholders: How does stakeholders' Facebook engagement help or ruin a company's reputation? Public Relat Rev 2017 Mar;43(1):201-210. [doi: 10.1016/j.pubrev.2016.12.004]
112. Leask J. Target the fence-sitters. Nature 2011 May 26;473(7348):443-445. [doi: 10.1038/473443a] [Medline: 21614055]
113. Roozenbeek J, Schneider CR, Dryhurst S, Kerr J, Freeman ALJ, Recchia G, et al. Susceptibility to misinformation about COVID-19 around the world. R Soc Open Sci 2020 Oct;7(10):201199 [FREE Full text] [doi: 10.1098/rsos.201199] [Medline: 33204475]

Abbreviations

ANOVA: analysis of variance
HPV: human papillomavirus
IFCN: International Fact-Checking Network

NB2: negative binomial regression
RQ: research question
VPD: vaccine-preventable disease
WHO: World Health Organization

Edited by M Gisondi, M Gottlieb; submitted 08.03.22; peer-reviewed by A Koon, L Tang; comments to author 19.04.22; revised version received 16.06.22; accepted 21.06.22; published 06.07.22.

Please cite as:
Ngai CSB, Singh RG, Yao L
Impact of COVID-19 Vaccine Misinformation on Social Media Virality: Content Analysis of Message Themes and Writing Strategies
J Med Internet Res 2022;24(7):e37806
URL: https://www.jmir.org/2022/7/e37806
doi:10.2196/37806
PMID:35731969

Original Paper

Chatbot-Delivered COVID-19 Vaccine Communication Message Preferences of Young Adults and Public Health Workers in Urban American Communities: Qualitative Study

Rose Weeks[1,2], MPH; Lyra Cooper[1], MHS; Pooja Sangha[1], MHS; João Sedoc[3,4], PhD; Sydney White[1], MPH; Assaf Toledo[5], PhD; Shai Gretz[5], MSc; Dan Lahav[5], MSc; Nina Martin[6], PhD; Alexandra Michel[1], MPH; Jae Hyoung Lee[1], MD; Noam Slonim[5], PhD; Naor Bar-Zeev[1], PhD

[1]International Vaccine Access Center, Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[2]Center for American Indian Health, Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[3]Stern School of Business, New York University, New York, NY, United States

[4]Whiting School of Engineering, Johns Hopkins University, Baltimore, MD, United States

[5]IBM Research AI, Haifa, Israel

[6]Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

Corresponding Author:
Rose Weeks, MPH
International Vaccine Access Center
Department of International Health
Johns Hopkins Bloomberg School of Public Health
415 N Washington Street 5th Floor
Baltimore, MD, 21231
United States
Phone: 1 3478046526
Email: rweeks@jhu.edu

Abstract

Background: Automated conversational agents, or chatbots, have a role in reinforcing evidence-based guidance delivered through other media and offer an accessible, individually tailored channel for public engagement. In early-to-mid 2021, young adults and minority populations disproportionately affected by COVID-19 in the United States were more likely to be hesitant toward COVID-19 vaccines, citing concerns regarding vaccine safety and effectiveness. Successful chatbot communication requires purposive understanding of user needs.

Objective: We aimed to review the acceptability of messages to be delivered by a chatbot named VIRA from Johns Hopkins University. The study investigated which message styles were preferred by young, urban-dwelling Americans as well as public health workers, since we anticipated that the chatbot would be used by the latter as a job aid.

Methods: We conducted 4 web-based focus groups with 20 racially and ethnically diverse young adults aged 18-28 years and public health workers aged 25-61 years living in or near eastern-US cities. We tested 6 message styles, asking participants to select a preferred response style for a chatbot answering common questions about COVID-19 vaccines. We transcribed, coded, and categorized emerging themes within the discussions of message content, style, and framing.

Results: Participants preferred messages that began with an empathetic reflection of a user concern and concluded with a straightforward, fact-supported response. Most participants disapproved of moralistic or reasoning-based appeals to get vaccinated, although public health workers felt that such strong statements appealing to communal responsibility were warranted. Responses tested with humor and testimonials did not appeal to the participants.

Conclusions: To foster credibility, chatbots targeting young people with vaccine-related messaging should aim to build rapport with users by deploying empathic, reflective statements, followed by direct and comprehensive responses to user queries. Further studies are needed to inform the appropriate use of user-customized testimonials and humor in the context of chatbot communication.

(J Med Internet Res 2022;24(7):e38418) doi:10.2196/38418

KEYWORDS

vaccine hesitancy; COVID-19; chatbots; AI; artificial intelligence; natural language processing; social media; vaccine communication; digital health; misinformation; infodemic; infodemiology; conversational agent; public health; user need; vaccination; health communication; online health information

Introduction

Vaccine hesitancy has emerged as a public health threat as trust in immunization systems has been strained across much of the world [1,2]. Global measles outbreaks occurring in the face of waning vaccine uptake propelled vaccine hesitancy onto the World Health Organization's list of top global health concerns [3,4]. The COVID-19 pandemic has brought hesitancy into sharp focus, including in the United States, which experienced one of the highest COVID-19 mortality rates among high-income nations [5].

A survey of more than 5 million Americans conducted via Facebook found that adults aged 18-34 years had the highest rates of vaccine hesitancy in May 2021 [6]. Moreover, despite disproportionately high COVID-19 mortality rates within communities of color [7], younger adults and Black, American Indian or Alaska Native, and multiracial groups continued to be the most hesitant, citing concerns regarding vaccine development, safety, and effectiveness [8-11]. As of May 2022, 3 in 10 Americans eligible for a COVID-19 vaccine have yet to get fully vaccinated, fueling continued disease spread and hindering pandemic recovery efforts [12].

To combat hesitancy, agency and advocacy leaders drew upon decades of communication science learning about building vaccine demand. Such guidance included the need for proactive planning efforts to understand the target populations, audience segmentation, tailored messaging, selection of appropriate channels, and commercial marketing approaches to ensure vaccines could be delivered via convenient and accessible services [13-16]. Program planning efforts would include continual analysis of the information landscape for competition, including misinformation and disinformation [15,17]. To build trust and engage young audiences often complacent about individual risk for disease, vaccine communication should be 2-way—listening and telling in equal measures—and in-person as well as web-based [13]. In urban communities, initiatives began by acknowledging historical injustices and ongoing inequities that drive distrust, with community-based health educators deployed to listen to concerns and provide support in person [18].

As a tool for providing tailored messaging, social listening, and 2-way dialogue, automated conversational agents, or chatbots, were cited early in the pandemic as a promising tool to offer COVID-19 health guidance on demand [19,20]. Chatbots have provided support on a range of health issues including chronic disease, addiction, reproductive health, depression, and anxiety, with promising adaptations of evidence-based interventions such as cognitive behavioral therapy [21-25]. This approach may appeal especially to young adults, since a substantial proportion of millennials, born from 1981-1996, are more trusting of web-based information and better equipped to use health technology than earlier generations [26,27]. Since the start of the pandemic, chatbots have been designed to provide COVID-19 health guidance in experimental settings [28], with some available globally via messaging platforms such as WhatsApp and Telegram [29,30]. Given their engaging, dialogue-based design, chatbots have a role in reinforcing public health guidance disseminated via other interventions such as social media campaigns. However, there is limited evidence related to the message design and framing of vaccine-related content delivered over digital platforms (eg, social media) and very little known about how such messaging should be delivered by educational chatbots in public health contexts [31].

Formative research has enabled the production of tailored content to optimize the delivery of messages [32]. An overarching factor in engaging and persuading audiences is the presence of credibility and trust, which can be defined as a combination of integrity, dependability, and competence [33,34]. Continually assessed by audiences, credibility can be lost through the delivery of a muddled or apparently dishonest message. Perhaps most central to vaccine communication in the context of hesitancy is the use of empathy, or a sense of one speaker understanding the experience of another. Empathic and reflective statements are critical components in motivational interviews, one of the few evidence-based means to soften vaccine hesitancy [35,36].

Seeking to review the acceptability of messages to be delivered by a chatbot, we engaged with potential users to identify which styles were preferred by young, urban-dwelling Americans. We also studied message reception with public health workers, anticipating that the chatbot would be used by the latter as a job aid. This formative research supported the development of a COVID-19 vaccine chatbot, VIRA, which was launched in 2021 by the International Vaccine Access Center at the Johns Hopkins Bloomberg School of Public Health [37]. IBM Research developed and managed the chatbot's back end, which used Key Point Analysis, a commercially available technology that facilitates "extractive summarization" to process numerous comments, opinions, and statements and reveal the most important points and their relative prevalence. VIRA was initially programmed to respond to 100 Key Points, with up to 4 styles of responses to each identified concern. Key Points or distinct vaccine concerns were identified through various means: conducting a Twitter analysis, reviewing audience questions in Zoom-based public forums hosted by our affiliated academic centers, and synthesizing web pages with frequently asked questions [38-40]. To draft responses, we considered previous evidence that emphasizing social and physical consequences in an emotional format elicits broad influence [14], as well as evidence that trust is established through the perceptions of care and concern [41,42]. VIRA's response database initially

consisted of factual-only responses (responses containing data-driven information), empathy-factual responses (factual responses beginning with an empathetic phrase that validated the user's query), principled responses (responses that appeal to a user's conscience, often referencing community well-being [43]), rational arguments (responses containing a logical argument), testimonials (responses featuring a quote from a reputable expert), and humorous responses. All responses sought to minimize technical language and word count (under 280 characters). In this analysis, we investigated the appropriateness and tailoring of these responses.

Methods

Recruitment

Through 4 semistructured focus group discussions (FGDs), we assessed the appropriateness of different styles of responses to common COVID-19 vaccine questions. We selected focus groups to generate insightful participant discussions to illustrate group norms and diversity in the sampled population within a short period of time [44]. We recruited 2 participant groups in the United States: (1) urban-dwelling individuals aged 18-28 years and (2) public health workers, defined as individuals contracted or employed by health departments to encourage the uptake of COVID-19 vaccines. To identify health workers, we used snowball sampling through professional contacts in urban health departments of the United States. We also posted ads on Craigslist and Twitter, targeting both health workers and young people in Baltimore, Charlotte, New York City, Philadelphia, and Washington, D.C. We aimed to achieve maximum variability in race and ethnicity for both participant groups to explore attitudes toward chatbots providing health information. Since the chatbot was aimed to support people along the continuum of vaccine hesitancy up to vaccine refusal, and our study aimed to encourage productive group discussions among individuals with some openness to change around vaccination, we excluded people who stated they would "definitely NOT choose to get a COVID-19 vaccine by August 2021" in a scaled response [45,46].

Ethics Approval

The Johns Hopkins Bloomberg School of Public Health Institutional Review Board approved this study (approval 15714).

Data Collection

Following individually obtained informed consent, participants completed a demographic questionnaire using Qualtrics software (SAP America). Each participant then joined a web-based FGD via Zoom (licensed account; Zoom Video Communications Inc), 1 of which was composed exclusively of health workers and the other 3 composed of young participants [47]. Discussions were facilitated by a doctoral-level researcher, a masters-level faculty member, and a trained graduate student. Each FGD lasted 1 hour and had a maximum of 8 participants. Facilitators introduced VIRA, a chatbot developed by Johns Hopkins University that provided answers to common COVID-19 vaccine questions. Participants viewed 7 cards containing a question related to COVID-19 vaccination and 3 or 4 proposed responses written in various message styles. Participants were asked to select their preferred response, and conversation was encouraged between participants to further explore preferences. Table 1 displays sample messages (see Multimedia Appendix 1 for all message content).

Table 1. Sample questions and tested response styles.

Question or comment, response type	Response
I'm not sure if the vaccine is safe, so I want to see how it affects others before I get vaccinated.	
Factual-only response	All vaccines go through clinical trials to test safety and effectiveness. For the COVID-19 vaccines, the FDA[a] set up rigorous standards for vaccine developers to meet and thousands of people worldwide participated in clinical trials before the vaccines became available to the public!
Empathy-factual response	This is an important question for many people! Once a vaccine is authorized for use, monitoring continues with systems in place to track problems or side-effects that were not detected during clinical trials. You can feel safe knowing these systems have got your back!
Principled response	It's very natural to have concerns. Yet, if some people choose to wait, we will not beat this pandemic any time soon. If you are willing to get vaccinated, you can do so knowing that millions have been safely vaccinated and you are helping our path to normalcy.
I'm worried about vaccine side effects and adverse reactions.	
Rational argument	The likelihood of experiencing a severe side-effect is very small—less than 5 out of 1,000 people! You'll probably just have some manageable side-effects that resolve in a few days.
I'm young and healthy, so why do I need to get vaccinated?	
Testimonial	A Harvard physician said, "while the vast majority of young adults who get COVID-19 are not going to require hospitalization, those who do have a really high risk for adverse outcomes." A vaccine can prevent severe illness, even if you're young and healthy.
Should I get the vaccine if I've had COVID-19?	
Humorous response	Spoiler: People who have COVID-19 should still get vaccinated, but only AFTER you get well!

[a]FDA: U.S. Food and Drug Administration.

Data Analysis

Recorded FGDs were transcribed using Temi software [48], with investigators reviewing each transcript for quality assurance. Dedoose software (version 9.0.46; SocioCultural Research Consultants) [49] was used for data management, such as coding, code report exports, and the reassembly process. We used a grounded theory approach to develop a conceptual framework of how people respond to the various styles of COVID-19 vaccine messaging that could inform the subsequent production of chatbot responses [50].

Through our inductive qualitative analysis process, we identified the emerging themes with which to code our data [51]. First, we developed our initial codebook, with 2 researchers independently reviewing and conducting line-by-line coding of 2 rich transcripts and producing open codes. The study team then convened to review and condense these open codes into broader themes and subthemes. Once the codebook was finalized, 2 researchers then coded each transcript and a third resolved any coding discrepancies. Although some coding redundancy was discovered, no new codes were identified, indicating a saturation of the themes outlined within the codebook [52]. Once the coding process was complete, the team arrayed the data into matrices to identify thematic patterns related to the code "chatbot credibility"—discussions of which were woven throughout FGDs, as seen in memos produced during coding and reassembly. To understand participants' overall preference for certain message styles, we also produced a count of preferences for each message type reviewed during the FGD process.

Results

Participant Characteristics

Between June and October 2021, several months after COVID-19 vaccines became widely available in the United States, we held 4 web-based focus groups with a total of 7 individuals aged 25-61 years working in public health or vaccine outreach roles and 13 people aged 18-28 years (see Table 2 for participant demographics). Of the 20 participants, most (80%, n=16) were women, with a mean age of 28.5 years. The median self-reported household income was US $56,000; for younger participants, this likely included parental income. Most (90%, n=18) participants said they were vaccinated against COVID-19.

Table 2. Participant demographics.

Characteristic	Participant
Self-identified gender (N=20), n (%)	
Female	16 (80)
Male	4 (20)
Highest level of education (N=20), n (%)	
High school graduate	4 (20)
Some college, no degree	3 (15)
Bachelor's degree or higher	13 (65)
Self-identified race or ethnicity (N=20), n (%)	
American Indian or Alaska Native	1 (5)
Asian or Pacific Islander	4 (20)
Black or African American	5 (25)
Hispanic or Latino	3 (15)
White	4 (20)
More than 1 race or ethnicity	3 (15)
Age group (years; N=20), n (%)	
18-29	15 (75)
30-49	3 (15)
50-69	2 (10)
Self-reported annual household income (2021; US $; n=19), median (range)	56,000 (25,000-200,000)
Vaccinated (N=20), n (%)	18 (90)

Thematic Findings

Analysis of FGDs identified themes describing the message preferences of young adults and public health workers in urban American communities. In the following quotes, we describe participants as either public health workers ("advocates") or young people.

Credibility

The credibility of a chatbot message was the predominant theme influencing response selection. Both young participants and

advocates said messages achieved credibility through (1) message directness and (2) the establishment of rapport between the user and chatbot through conversational syntax and empathetic, reflective statements. Of the 26 total responses, 6 (23%) consisted of an empathy-factual composite style, beginning with a user-centered, reflective message such as "it sounds like you have concerns" or "this is an important question for a lot of people." Empathetic responses used casual, nontechnical language to answer questions using evidence in what participants said was a transparent, contextualized response. A young woman described the style as, "kind of sticking to the facts in a colloquial/conversational manner—doesn't feel like I'm reading a newspaper or a research paper" (Participant 09).

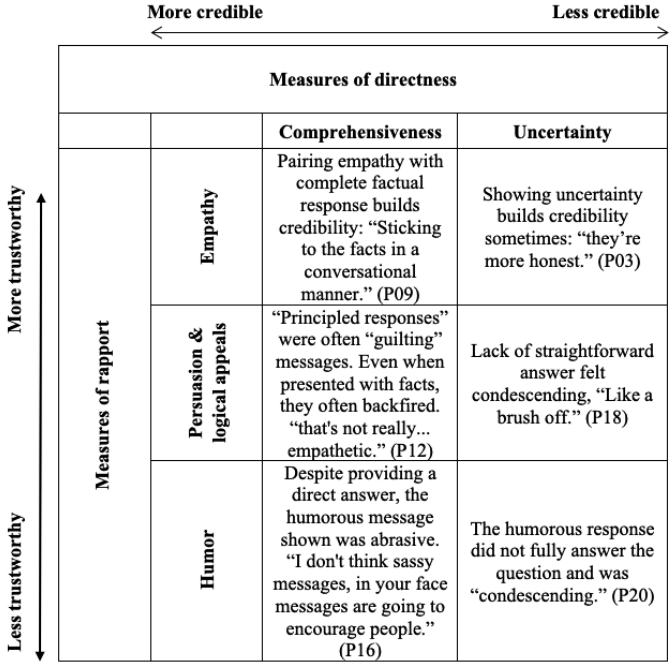
Another young woman liked the conversational tone and "extensive" detail provided in a factual-style message about side effects, which stated that the vaccine's side effects "should resolve within one or two days of vaccination." She said, "It covered it pretty extensively. It sets me up for what I can expect. And then I would feel more secure knowing...the chatbot, it's giving me [a] correct answer" (Participant 19).

Figure 1 illustrates how messages achieved credibility through rapport-building and directness and how participants felt messages lost credibility when responses didn't answer a question directly—"like a brush off." As seen in textual excerpts in the table, trustworthiness was eroded when messages compromised rapport, either by incorporating humor or by deploying guilt-laden arguments.

Both groups of participants regarded scientific messages as credible, saying they trusted the message since it was communicated by a Johns Hopkins University chatbot. In a typical response about how the brand affected message reception, a young man said: "I felt this [was] trustworthy, [be]cause I knew it came from like Johns Hopkins, which has a strong reputation" (Participant 07).

Although advocates and young participants both preferred empathetic statements prefacing a full, factual response, some felt such messages seemed inauthentic. The phrase "having doubts is normal," in the words of a male advocate, "makes [the chatbot] more humane, more human-like, and more accepting" (Participant 08). Meanwhile, participants felt the phrase "I hear you" was marginally reassuring, but the statement "I wondered about that too" seemed "weird and fake [from a chatbot]" to one young woman (Participant 05).

Figure 1. Message attributes supporting and hindering credibility with young focus group participants. Textual excerpts coded with both directness- and rapport-related variables (eg, each cell shows a textual passage double coded with a directness-related variable and rapport-related variable). P: participant.



Responsibility

Principled responses sought to persuade users to get vaccinated by appealing to an individual's responsibility for the well-being of a larger community and using reasoning instead of evidence (see example in Table 1). For instance, instead of stating that vaccines are safe because "All vaccines go through clinical trials" as in a factual response, a principled response would say, "If some people chose to wait, we will not beat this pandemic any time soon." Young participants saw this as evasive as well as "condescending" and "aggressive." In the words of several participants, these messages were judgmental or shaming; as told by a young participant "You're telling them that you put your families, your community at risk too. You're making them feel guilty...I feel like that's not really...empathetic" (Participant 12).

However, the style had some appeal with 4 young participants of color. One participant, a young man, felt such direct messaging was warranted:

> *I like [principled response] D because at this point in time, I think we need more aggressive messaging. Like, guilt people, shake people, let's be serious...don't put your friends and family at risk. Just get the shot.* [Participant 10]

Meanwhile, advocates aged >30 years often preferred such messages, sometimes wanting messaging to be "stronger" and "louder" to combat misinformation around vaccine myths. As one advocate said, "I think that message should be really pushed out a little louder. [The vaccine] prevents you from getting deathly ill" (Participant 16).

Advocates preferred principled responses that emphasized shared responsibility to prevent COVID-19 spread. As one young advocate said, these responses "made me think about the risks I posed not just [for] myself, but the people around me" (Participant 01). Moreover, advocates shared concerns about their family members and discussed feeling surrounded by people that "weren't doing their part." As one advocate said, "We've been getting hit hard, especially in the Black and Indigenous, Latinx, API communities, and this thing isn't going anywhere anytime soon" (Participant 18).

An advocate recalled seeing community members previously hospitalized with COVID-19 "still not wearing your mask...it just made me more weary" (Participant 17). Such fatigue with community members not taking precautions to protect themselves and one another was linked with participant preference for principled messaging that was direct and insistent on communal responsibilities.

Resistance to Logical Appeals

Participants rarely preferred the rational arguments shown. We tested the following message in response to the question, "Are COVID vaccines worse than the disease itself?"

> *The trouble with that logic is that it's difficult to predict who will survive an infection without becoming a COVID-19 long hauler. Almost 30% of people who've survived COVID-19 still experience long-term side-effects!*

One participant commented that the tone of the response sounded "judgmental." Similarly, a young woman said "it was like the most convincing argument as to why you would want to get the vaccine because like it shows how many people get long-term side effects, but I did agree that the tone...was a little condescending" (Participant 20).

Humor

Almost unanimously, participants disapproved of the humorous response shown and said it mocked people for asking questions. Both young people and advocates said it was "pushy," "saucy," and "condescending." Moreover, participants said it did not fully answer the question or provide context to support statements.

Balancing Comprehensiveness and Uncertainty

After "comprehensiveness," the code "credibility" was most likely to overlap with "transparency," indicating the importance of directly answering a question without seeming to withhold information. Although both advocates and young people discussed wanting responses to be both direct and comprehensive, the participants did not agree about explicitly highlighted scientific uncertainty. For instance, one message said, "scientists aren't totally sure" about whether vaccines stop transmission. Advocates said acknowledging ongoing studies was appropriate, since "we're still learning about it," in the words of an advocate (Participant 18). However, young participants disagreed, saying phrases acknowledging uncertainty were unsettling and did not promote vaccination, with a young participant stating "she [the chatbot] seemed super uncertain" or "it almost makes it seem like people should wait to see more studies [to get the vaccine]" (Participant 19).

Authority as Elitism

Young participants considered the use of a testimonial-style quote attributed to a Harvard physician to be elitist. One male advocate responded by saying, "Why do I care? It's throw[ing] that he just has a title at my face" (Participant 08). Advocates aged >30 years agreed that the testimonial was not helpful, citing the politicization of doctors and science and suggesting the chatbot display testimonials from frontline health workers, such as emergency medical technicians.

Relative Message Preference

To triangulate and strengthen our qualitative analysis, we tallied the number of votes the participants cast indicating their preferences for the messages on each of the 7 cards shown. FGD participants voted for a preferred message a total of 84 times (some did not select a response for each card shown). Participants voted for empathetic-factual messages 40 (48%) times, over 50% more times than factual-only messages—which at 24 (29%) votes was the second most preferred message style. However, although young participants most often (51%, 31/61) selected the empathetic-factual messages presented on message cards, public health workers most often (38%, 8/21) selected a principled response; young people rarely (8%, 5/61) selected this style with a few exceptions described above. Participants infrequently (8%, 7/84) preferred rational arguments, and participants never selected the testimonial or humorous messages, although just 1 example of each were shown on the cards. Although these quantitative results are not statistically significant given the qualitative study design and small sample size, the overwhelming preference for empathetic-framed responses among young participants is notable.

Discussion

Principal Findings

In this formative study of preferences for messages delivered by a COVID-19 chatbot, participants from urban American communities favored messages that were empathetic, direct, and comprehensive in answering questions related to COVID-19 vaccination. Messages achieved credibility through a combination of empathy and straightforward, evidence-based responses. User-centered reflective statements and conversational language that minimized the use of technical jargon fostered rapport between the chatbot and user.

Public health messages often contain statistics and rational statements, a strategy shown to effectively counter vaccine misinformation [53]. However, among our participants, most of whom were already vaccinated, this strategy alone was not as successful as messages that also included empathetic statements. Other studies among Black Americans with chronic conditions during COVID-19 found that participants also preferred chatbots to be "personable and empathetic" [54]. Other empathy-simulating chatbots have reported similar positive feedback from users [55-57]. Empathetic statements validated people's search for knowledge and implicitly acknowledged the loneliness of the pandemic experience [58].

Participants in our study preferred straightforward, comprehensive responses that are similar to answers from an informed and respectful human interlocutor. The chatbot needed to completely answer user questions, or else may be perceived as evasive and potentially untrustworthy. Such expectations for politeness and respect align with the observations of technological anthropomorphizing, building on studies that show individuals' interactions with computers are "fundamentally social" and that people naturally characterize the computer as a social actor [59-61].

For most young study participants, principled responses—messages appealing to concerns for family and community—counteracted the chatbot's attempts to simulate empathy. In contrast, a meta-analysis of 60 studies found that younger audiences were the group most influenced by messages highlighting social consequences, showing the complexity of parsing message tactics in vaccine science when layered on top of a pandemic context [62]. However, for public health workers and several young people of color, the strong appeal to solidarity resonated with pandemic fatigue and frustration with individuals remaining unvaccinated in the face of widespread community distress; this finding is reinforced by the meta-analysis, which showed that so-called high-involvement audiences prefer data and strong messages that are somewhat fearful [62]. Built with audience-tested messages, the chatbot could offer a framework to support communication between health workers and community members that would integrate facilitated empathy.

Similarly, "rational arguments" eroded rapport between participants and the chatbot. Social media–related studies have used similar framing to "inoculate" audiences against misinformation, but this approach was not well-received in our study in the context of a chatbot [63,64].

The testimonial message shown to participants in the study was unappealing because the spokesperson was viewed as elitist. Future testimonials used in chatbot messages could be matched to participant demographics [65].

Although the humorous message shown to participants in this study performed poorly, humor is increasingly used to reach social media audiences otherwise disengaged from a public health topic and promote widespread sharing [66-68]. Friendly, self-deprecating humor as seen in popular voice assistant bots may be a better choice for one-to-one anonymous conversations with a chatbot than meme-style humor [69].

Limitations

Although this study revealed COVID-19 vaccine message preferences for chatbot conversations with young, urban community members in the United States, it has several limitations. First, our participants were mostly college-educated. Given our reliance on Twitter ads for recruitment, this may be because Twitter users tend to be better educated and more left-leaning [70]. In addition, our chatbot's institutional branding, widely known to promote pandemic mitigation measures, likely dissuaded some vaccine skeptics. Despite the fact that most participants were vaccinated, a range of concerns regarding COVID-19 vaccines was proffered, and we regard hesitancy as a reflection of a wide spectrum of concerns and views, including among those deciding to get vaccinated. The uncertainties of the pandemic were challenging to our study's feasibility. Since we needed to collect data rapidly to iteratively redesign a tool already in use by public health departments, we held only 4 focus groups, limiting the potential transferability of the findings to similar populations in other geographic areas or among other communities. However, due to the observed permeability of the US population segments of vaccinated and unvaccinated people, we believe the results are relevant to support efforts to counteract vaccine hesitancy [6]. The pandemic was a highly dynamic environment for studying a tool to counter vaccine hesitancy, and the well-publicized, highly contagious delta variant spread was concurrent with recruitment, increasing the uptake of vaccines; in the subsequent months, residual concerns in vaccinated members of the public have surfaced as well as the reluctance to get booster doses [71]. Further, most participants identified with or worked in communities with high rates of vaccine hesitancy and referenced the concerns of community members.

Moreover, preferences may not predict web-based behavior, and the appraisal of messages in the FGDs may diverge from assessments in the context of a dialogue-based chatbot [72]; additionally, due to the constraints of the hour-long discussions, we were unable to show multiple varieties of humor- and testimonial-style messages that may have yielded different findings. Other determinants of message acceptance, including the influence of gain- or loss-framing on chatbot message preference, could be further explored, and ultimately, the chatbot's impact on actual behavior should be evaluated [62,73].

Conclusions

This study focused on the establishment of credible messaging from a chatbot to be used by young Americans and public health workers during the peak of the COVID-19 pandemic. We found that for young audiences, message credibility was optimized with empathetic statements and comprehensive, direct, and evidence-rich content. Pandemic-weary advocates and some young people from communities disproportionately affected by COVID-19 tended to prefer stronger, responsibility-focused messaging. Although credibility is essential to persuading users of a position, persuasion was not the target of this study. Additional controlled and randomized studies are needed to determine if a chatbot could persuade users to change their perception of the safety and effectiveness of vaccines and get vaccinated.

Acknowledgments

This research and the development of the chatbot VIRA [37] was supported by the Johns Hopkins COVID-19 Training Initiative with funding from Bloomberg Philanthropies and in-kind support from IBM Research.

Conflicts of Interest

NB-Z received research grants from Johnson & Johnson and Merck for unrelated work outside the scope of this paper. All authors declare no other conflicts of interest.

Multimedia Appendix 1
Full range of messages shown to focus group participants.
[DOCX File , 22 KB - jmir_v24i7e38418_app1.docx]

References

1. Ozawa S, Stack ML. Public trust and vaccine acceptance--international perspectives. Hum Vaccin Immunother 2013 Aug;9(8):1774-1778 [FREE Full text] [doi: 10.4161/hv.24961] [Medline: 23733039]
2. Gross L. A broken trust: lessons from the vaccine--autism wars. PLoS Biol 2009 May 26;7(5):e1000114 [FREE Full text] [doi: 10.1371/journal.pbio.1000114] [Medline: 19478850]
3. Worldwide measles deaths climb 50% from 2016 to 2019 claiming over 207 500 lives in 2019. World Health Organization. 2020 Nov 12. URL: https://tinyurl.com/2jbc6he5 [accessed 2022-05-16]
4. Ten threats to global health in 2019. World Health Organization. URL: https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019 [accessed 2022-05-23]
5. Mortality analyses. Johns Hopkins Coronavirus Resource Center. URL: https://coronavirus.jhu.edu/data/mortality [accessed 2022-05-16]
6. King WC, Rubinstein M, Reinhart A, Mejia R. Time trends, factors associated with, and reasons for COVID-19 vaccine hesitancy: a massive online survey of US adults from January-May 2021. PLoS One 2021 Dec 21;16(12):e0260731 [FREE Full text] [doi: 10.1371/journal.pone.0260731] [Medline: 34932583]
7. Siegel M, Critchfield-Jain I, Boykin M, Owens A. Actual racial/ethnic disparities in COVID-19 mortality for the non-Hispanic Black compared to non-Hispanic White population in 35 US states and their association with structural racism. J Racial Ethn Health Disparities 2022 Jun;9(3):886-898 [FREE Full text] [doi: 10.1007/s40615-021-01028-1] [Medline: 33905110]
8. Callaghan T, Moghtaderi A, Lueck JA, Hotez P, Strych U, Dor A, et al. Correlates and disparities of intention to vaccinate against COVID-19. Soc Sci Med 2021 Mar;272:113638 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113638] [Medline: 33414032]
9. Fisher KA, Bloomstone SJ, Walder J, Crawford S, Fouayzi H, Mazor KM. Attitudes toward a potential SARS-CoV-2 vaccine : a survey of U.S. adults. Ann Intern Med 2020 Dec 15;173(12):964-973 [FREE Full text] [doi: 10.7326/M20-3569] [Medline: 32886525]
10. Ruiz JB, Bell RA. Predictors of intention to vaccinate against COVID-19: results of a nationwide survey. Vaccine 2021 Feb 12;39(7):1080-1086 [FREE Full text] [doi: 10.1016/j.vaccine.2021.01.010] [Medline: 33461833]
11. Khubchandani J, Macias Y. COVID-19 vaccination hesitancy in Hispanics and African-Americans: a review and recommendations for practice. Brain Behav Immun Health 2021 Aug;15:100277 [FREE Full text] [doi: 10.1016/j.bbih.2021.100277] [Medline: 34036287]
12. COVID data tracker. Centers for Disease Control and Prevention. URL: https://covid.cdc.gov/covid-data-tracker [accessed 2022-02-27]
13. Goldstein S, MacDonald NE, Guirguis S, SAGE Working Group on Vaccine Hesitancy. Health communication and vaccine hesitancy. Vaccine 2015 Aug 14;33(34):4212-4214 [FREE Full text] [doi: 10.1016/j.vaccine.2015.04.042] [Medline: 25896382]
14. Nowak GJ, Gellin BG, MacDonald NE, Butler R, SAGE Working Group on Vaccine Hesitancy. Addressing vaccine hesitancy: the potential value of commercial and social marketing principles and practices. Vaccine 2015 Aug 14;33(34):4204-4211 [FREE Full text] [doi: 10.1016/j.vaccine.2015.04.039] [Medline: 25900132]
15. Evans WD, French J. Demand creation for COVID-19 vaccination: overcoming vaccine hesitancy through social marketing. Vaccines (Basel) 2021 Apr 01;9(4):319 [FREE Full text] [doi: 10.3390/vaccines9040319] [Medline: 33915695]
16. Butler R, MacDonald NE, SAGE Working Group on Vaccine Hesitancy. Diagnosing the determinants of vaccine hesitancy in specific subgroups: the guide to Tailoring Immunization Programmes (TIP). Vaccine 2015 Aug 14;33(34):4176-4179 [FREE Full text] [doi: 10.1016/j.vaccine.2015.04.038] [Medline: 25896376]
17. Vaccine misinformation management field guide. United Nations Children's Fund (UNICEF). 2020 Dec. URL: https://vaccinemisinformation.guide [accessed 2021-02-02]

18. Strully KW, Harrison TM, Pardo TA, Carleo-Evangelist J. Strategies to address COVID-19 vaccine hesitancy and mitigate health disparities in minority populations. Front Public Health 2021 Apr 23;9:645268 [FREE Full text] [doi: 10.3389/fpubh.2021.645268] [Medline: 33968884]
19. Miner AS, Laranjo L, Kocaballi AB. Chatbots in the fight against the COVID-19 pandemic. NPJ Digit Med 2020 May 04;3:65 [FREE Full text] [doi: 10.1038/s41746-020-0280-0] [Medline: 32377576]
20. Amiri P, Karahanna E. Chatbot use cases in the Covid-19 public health response. J Am Med Inform Assoc 2022 Apr 13;29(5):1000-1010 [FREE Full text] [doi: 10.1093/jamia/ocac014] [Medline: 35137107]
21. Bonnevie E, Lloyd TD, Rosenberg SD, Williams K, Goldbarg J, Smyser J. Layla's Got You: developing a tailored contraception chatbot for Black and Hispanic young women. Health Educ J 2021;80(4):413-424. [doi: 10.1177/0017896920981122]
22. 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 17;23(2):e25060 [FREE Full text] [doi: 10.2196/25060] [Medline: 33484114]
23. Pereira J, Díaz Ó. Using health chatbots for behavior change: a mapping study. J Med Syst 2019 Apr 04;43(5):135. [doi: 10.1007/s10916-019-1237-1] [Medline: 30949846]
24. Wysa. URL: https://www.wysa.io/ [accessed 2021-02-16]
25. Copper-Ind C. Aetna partners with well-being app to provide mental healthcare during pandemic. International Investment. 2020 May 19. URL: https://www.internationalinvestment.net/news/4015424/aetna-partners-app-provide-mental-healthcare-pandemic [accessed 2021-02-16]
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. Journal of Retailing and Consumer Services 2020 Nov;57:102221. [doi: 10.1016/j.jretconser.2020.102221]
27. Yasgur BS. Millennials flock to telehealth, online research. WebMD. 2021 Apr 02. URL: https://www.webmd.com/lung/news/20210402/millennials-flock-to-telehealth-online-research [accessed 2022-03-30]
28. Altay S, Hacquin AS, Chevallier C, Mercier H. Information delivered by a chatbot has a positive impact on COVID-19 vaccines attitudes and intentions. J Exp Psychol Appl Preprint posted on October 28, 2021. [doi: 10.1037/xap0000400] [Medline: 34726454]
29. WHO Health Alert brings COVID-19 facts to billions via WhatsApp. World Health Organization. 2021 Apr 26. URL: https://www.who.int/news-room/feature-stories/detail/who-health-alert-brings-covid-19-facts-to-billions-via-whatsapp [accessed 2021-07-26]
30. Almalki M, Azeez F. Health chatbots for fighting COVID-19: a scoping review. Acta Inform Med 2020 Dec;28(4):241-247 [FREE Full text] [doi: 10.5455/aim.2020.28.241-247] [Medline: 33627924]
31. Limaye RJ, Holroyd TA, Blunt M, Jamison AF, Sauer M, Weeks R, et al. Social media strategies to affect vaccine acceptance: a systematic literature review. Expert Rev Vaccines 2021 Aug;20(8):959-973. [doi: 10.1080/14760584.2021.1949292] [Medline: 34192985]
32. Noar SM. A 10-year retrospective of research in health mass media campaigns: where do we go from here? J Health Commun 2006;11(1):21-42. [doi: 10.1080/10810730500461059] [Medline: 16546917]
33. National Academies of Sciences, Engineering, and Medicine. Communicating Science Effectively: A Research Agenda. Washington, DC: The National Academies Press; 2017.
34. Renn O, Levine D. Credibility and trust in risk communication. In: Kasperson RE, Stallen PJM, editors. Communicating Risks to the Public. Dordrecht, Netherlands: Springer; 2010:175-217.
35. Cole JW, Chen AMH, McGuire K, Berman S, Gardner J, Teegala Y. Motivational interviewing and vaccine acceptance in children: The MOTIVE study. Vaccine 2022 Mar 15;40(12):1846-1854. [doi: 10.1016/j.vaccine.2022.01.058] [Medline: 35153096]
36. Gagneur A, Lemaître T, Gosselin V, Farrands A, Carrier N, Petit G, et al. A postpartum vaccination promotion intervention using motivational interviewing techniques improves short-term vaccine coverage: PromoVac study. BMC Public Health 2018 Jun 28;18(1):811 [FREE Full text] [doi: 10.1186/s12889-018-5724-y] [Medline: 29954370]
37. Vira, the Vax Chatbot. URL: https://vaxchat.org/ [accessed 2022-03-29]
38. About COVID-19 materials for tribes. Center for American Indian Health. URL: https://caih.jhu.edu/home/about-covid-19-materials-for-tribes [accessed 2022-03-05]
39. Answers to all your questions about getting vaccinated for Covid-19. The New York Times. 2021 Oct 18. URL: https://www.nytimes.com/interactive/2021/well/covid-vaccine-questions.html [accessed 2022-05-10]
40. Twitter API. Twitter. URL: https://developer.twitter.com/en/docs/twitter-api [accessed 2022-05-10]
41. Chen X, Hay JL, Waters EA, Kiviniemi MT, Biddle C, Schofield E, et al. Health literacy and use and trust in health information. J Health Commun 2018;23(8):724-734 [FREE Full text] [doi: 10.1080/10810730.2018.1511658] [Medline: 30160641]
42. Peters RG, Covello VT, McCallum DB. The determinants of trust and credibility in environmental risk communication: an empirical study. Risk Anal 1997 Feb;17(1):43-54. [doi: 10.1111/j.1539-6924.1997.tb00842.x] [Medline: 9131825]

43. Bilu Y, Gera A, Hershcovich D, Sznajder B, Lahav D, Moshkowich G, et al. Argument invention from first principles. Proc 57th Annu Meet Assoc Comput Linguist Florence, Italy: Association for Computational Linguistics; 2019 Jul Presented at: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics; July 28 to August 2, 2019; Florence, Italy p. 1013-1026. [doi: 10.18653/v1/p19-1097]
44. Mack N, Woodsong C, MacQueen K, Guest G, Namey E. Qualitative Research Methods: A Data Collector's Field Guide. Research Triangle Park, NC: Family Health International; 2005.
45. Report of the SAGE Working Group on vaccine hesitancy. World Health Organization. 2014 Nov 12. URL: https://tinyurl.com/2p8uk64j [accessed 2022-06-24]
46. 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]
47. Video conferencing, cloud phone, webinars, chat, virtual events. Zoom. URL: https://zoom.us/ [accessed 2022-03-29]
48. Temi. URL: https://www.temi.com/ [accessed 2022-03-27]
49. User guide. Dedoose. URL: https://www.dedoose.com/userguide/appendix [accessed 2022-02-28]
50. Corbin J, Strauss A. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. 4th ed. Thousand Oaks, CA: SAGE Publications; 2008.
51. Charmaz K. Constructing Grounded Theory. 2nd ed. Thousand Oaks, CA: SAGE Publications; 2014.
52. Hennink MM, Kaiser BN, Weber MB. What influences saturation? estimating sample sizes in focus group research. Qual Health Res 2019 Aug;29(10):1483-1496 [FREE Full text] [doi: 10.1177/1049732318821692] [Medline: 30628545]
53. Vraga EK, Kim SC, Cook J. Testing logic-based and humor-based corrections for science, health, and political misinformation on social media. J Broadcast Electron Media 2019 Sep 20;63(3):393-414. [doi: 10.1080/08838151.2019.1653102]
54. Kim J, Muhic J, Park S, Robert L. Trustworthy conversational agent design for African Americans with chronic conditions during COVID-19. In: Realizing AI in Healthcare: Challenges Appearing in the Wild. 2020 Presented at: 2021 CHI Conference on Human Factors in Computing Systems (CHI 2021); May 8-13, 2021; Online Virtual Conference (originally Yokohama, Japan). [doi: 10.7302/277]
55. Ta V, Griffith C, Boatfield C, Wang X, Civitello M, Bader H, et al. User experiences of social support from companion chatbots in everyday contexts: thematic analysis. J Med Internet Res 2020 Mar 06;22(3):e16235 [FREE Full text] [doi: 10.2196/16235] [Medline: 32141837]
56. Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. JMIR mHealth uHealth 2018 Nov 23;6(11):e12106 [FREE Full text] [doi: 10.2196/12106] [Medline: 30470676]
57. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
58. Czeisler M, Lane RI, Petrosky E, Wiley JF, Christensen A, Njai R, et al. Mental health, substance use, and suicidal ideation during the COVID-19 pandemic - United States, June 24-30, 2020. MMWR Morb Mortal Wkly Rep 2020 Aug 14;69(32):1049-1057 [FREE Full text] [doi: 10.15585/mmwr.mm6932a1] [Medline: 32790653]
59. Nass C, Steuer J, Siminoff ER. Computers are social actors. 1994 Apr 28 Presented at: CHI '94: Conference Companion on Human Factors in Computing Systems; April 24-28, 1994; Boston, MA p. 72-78. [doi: 10.1145/259963.260288]
60. Nass C, Moon Y. Machines and mindlessness: social responses to computers. J Social Isssues 2000 Jan;56(1):81-103. [doi: 10.1111/0022-4537.00153]
61. Christoforakos L, Feicht N, Hinkofer S, Löscher A, Schlegl SF, Diefenbach S. Connect With Me. Exploring Influencing Factors in a Human-Technology Relationship Based on Regular Chatbot Use. Front Digit Health 2021;3:689999 [FREE Full text] [doi: 10.3389/fdgth.2021.689999] [Medline: 34870266]
62. Keller PA, Lehmann DR. Designing effective health communications: a meta-analysis. Journal of Public Policy & Marketing 2008 Sep 01;27(2):117-130. [doi: 10.1509/jppm.27.2.117]
63. 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]
64. Cook J, Lewandowsky S, Ecker UKH. Neutralizing misinformation through inoculation: exposing misleading argumentation techniques reduces their influence. PLoS One 2017 May 5;12(5):e0175799 [FREE Full text] [doi: 10.1371/journal.pone.0175799] [Medline: 28475576]
65. Ramirez A, Despres C, Chalela P, Weis J, Sukumaran P, Munoz E, et al. Pilot study of peer modeling with psychological inoculation to promote coronavirus vaccination. Health Educ Res 2022 Mar 23;37(1):1-6. [doi: 10.1093/her/cyab042] [Medline: 35048119]
66. Public Health Messaging Can Be Funny Too!. American College Health Association. URL: https://www.youtube.com/watch?v=C_dgwhWhwGQ [accessed 2022-06-28]
67. Chen S, Forster S, Yang J, Yu F, Jiao L, Gates J, et al. Animated, video entertainment-education to improve vaccine confidence globally during the COVID-19 pandemic: an online randomized controlled experiment with 24,000 participants. Trials 2022 Feb 19;23(1):161 [FREE Full text] [doi: 10.1186/s13063-022-06067-5] [Medline: 35183238]

68. Vaala S, Ritter MB, Palakshappa D. Experimental effects of tweets encouraging social distancing: effects of source, emotional appeal, and political ideology on emotion, threat, and efficacy. J Public Health Manag Pract 2022 Apr;28(2):E586-E594. [doi: 10.1097/PHH.0000000000001427] [Medline: 34508051]
69. The AI companion who cares. Replika. URL: https://replika.ai [accessed 2021-07-26]
70. Wojcik S, Hughes A. Sizing up Twitter users. Pew Research Center. 2019 Apr 24. URL: https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/ [accessed 2022-05-10]
71. KFF COVID-19 vaccine monitor. Kaiser Family Foundation. URL: https://www.kff.org/coronavirus-covid-19/dashboard/kff-covid-19-vaccine-monitor-dashboard/ [accessed 2022-05-22]
72. Loft LH, Pedersen EA, Jacobsen SU, Søborg B, Bigaard J. Using Facebook to increase coverage of HPV vaccination among Danish girls: an assessment of a Danish social media campaign. Vaccine 2020 Jun 26;38(31):4901-4908. [doi: 10.1016/j.vaccine.2020.04.032] [Medline: 32362529]
73. Nan X. Communicating to young adults about HPV vaccination: consideration of message framing, motivation, and gender. Health Commun 2012;27(1):10-18. [doi: 10.1080/10410236.2011.567447] [Medline: 22276999]

Abbreviations

FGD: focus group discussion

Edited by M Gisondi, L Westafer; submitted 31.03.22; peer-reviewed by A Ruco, M Bardus; comments to author 04.05.22; revised version received 25.05.22; accepted 14.06.22; published 06.07.22.

Please cite as:
Weeks R, Cooper L, Sangha P, Sedoc J, White S, Toledo A, Gretz S, Lahav D, Martin N, Michel A, Lee JH, Slonim N, Bar-Zeev N Chatbot-Delivered COVID-19 Vaccine Communication Message Preferences of Young Adults and Public Health Workers in Urban American Communities: Qualitative Study
J Med Internet Res 2022;24(7):e38418
URL: https://www.jmir.org/2022/7/e38418
doi:10.2196/38418
PMID:35737898

Original Paper

Health Misinformation Across Multiple Digital Ecologies: Qualitative Study of Data From Interviews With International Students

Rashika Bahl[1], MIS; Shanton Chang[1], PhD; Dana McKay[2], PhD; George Buchanan[1], PhD

[1]School of Computing and Information Systems, University of Melbourne, Parkville, Australia

[2]School of Computing Technologies, RMIT University, Melbourne, Australia

Corresponding Author:
Rashika Bahl, MIS
School of Computing and Information Systems
University of Melbourne
Grattan Street
Parkville, 3053
Australia
Phone: 61 83441583
Email: rashikabahl@gmail.com

Abstract

Background: Transient migrants such as international students have received limited support from host country governments throughout the COVID-19 pandemic. An increase in misinformation, resulting in poor health outcomes for individuals, may impact an already vulnerable group.

Objective: Existing research examines the spread of misinformation. Similarly, there is extensive literature on the health information behavior of international students. However, there is a gap in the literature focusing on international students' interaction with health misinformation. This exploratory research aims to address this gap by examining international students' interaction with health misinformation during the COVID-19 pandemic.

Methods: A total of 11 participants took part in semistructured interviews and a health misinformation-identification exercise via Zoom. The data collected were subjected to qualitative thematic analysis. Multiple rounds of coding, checked by other coders, revealed 2 themes and 6 subthemes.

Results: The 2 main themes that emerged were (1) approaches to dealing with health misinformation and (2) how international students navigate across multiple digital ecologies. Results show that international students who draw on multiple digital ecologies for information reliably identify misinformation, suggesting that the use of multiple digital ecologies may have a protective effect against health misinformation.

Conclusions: Findings show that international students encounter health misinformation across multiple digital ecologies, and they also compare information across multiple ecologies. This comparison may support them in identifying health misinformation. Thus, the findings of this study combat narratives of international students' susceptibility to misinformation.

(J Med Internet Res 2022;24(7):e38523) doi:10.2196/38523

KEYWORDS

international students; social media; COVID-19; misinformation; infodemic; digital ecology; health information; information seeking; web-based health information

Introduction

During the COVID-19 pandemic, there has been concern about the vulnerability of migrant groups. Transient migrants such as international students, who have temporarily migrated to another country to study [1], have received limited support from the host country governments [2]. Moreover, an increase in misinformation on social media during the COVID-19 pandemic has contributed to vaccine hesitancy among migrant groups, resulting in poor health outcomes for individuals [3]. International students, due to their transient status, may draw on multiple digital ecologies to seek out information [4,5]. Drawing on multiple digital ecologies may influence international students' reactions to misinformation. They may experience an increased exposure to misinformation or miss out on crucial local information [6].

Existing literature covers the spread of misinformation [7-9] as well as the health information behavior of international students [10-12]. However, there is a gap in literature focusing on international students' interaction with health misinformation. This paper aims to address this gap in literature by examining the following question: "How do international students who use multiple digital ecologies interact with COVID-19 misinformation?" To do so, this paper will first synthesize the literature on the spread of misinformation and health information behavior of international students. Next, it will describe the methodology adopted. Finally, it will present results and discuss findings in the context of the literature.

Defining Misinformation and Social Media

"Misinformation" has a variety of synonyms in the literature, including fake news, spam, and trolling [13]. In this paper, similar to the study by Wu et al [13], we use "misinformation" in its broadest sense, including any false information regardless of the source or intent.

Social media, comprised of multiple information ecologies, refers to any internet-based or digital space where individuals can form and maintain connections as well as gather and share new information [14]. Information ecologies are systems made up of the individuals using the platforms, the technologies facilitating the platforms, and the values that drive the use of the platforms [7].

International Students and Health Misinformation

International students may be passive users of their host country health care systems due to language barriers and possessing limited information about their new host countries [11]. COVID-19 has resulted in a transition to telehealth worldwide [15], meaning that new arrivals must learn how to navigate health care systems digitally. International students have been shown to experience disconnectedness from their host countries, and digital spaces offer a line of connection to their communities [16]. Thus, international students often transcend digital ecologies as they interact with different groups of people in different geographic locations and with different cultural values [6]. Furthermore, COVID-19 policies vary across countries, which may influence international students' knowledge and perception of host country policies [12].

When navigating new digital ecologies, international students may turn to social media to gather health information from personal connections and have "guidance in clinical decision-making" [6,10]. They may also draw on their home country media sources for health information in their first language [10,11]. While moving through new digital ecologies for which they have limited context, international students rely on the limited information available on social media platforms to determine reliability [6]. If the international student is unaware that they are encountering misinformation, their experience will be functionally comparable to when they are encountering real information [6]; that is, if the misinformation resonates with them, they may acquire (mis)information [17,18]—information that is actually misinformation.

During the COVID-19 pandemic, there has been an increase in misinformation [19]. A study done on a sample of tweets related to COVID-19 showed that 24.8% of tweets contained misinformation, and 17.4% had unverifiable information [19]. International students who transcend digital ecologies may experience a doubling of misinformation across different digital ecologies. This can be damaging, as research shows if people view a story enough times, they are more likely to believe it, even if the story is being disproved [20]. Furthermore, research has demonstrated that context influences how misinformation is interpreted [21]. International students may encounter the same misinformation across digital ecologies. What might be framed accurately in one ecology could be framed inaccurately or in a biased manner in another. Biased information is a form of misinformation [13]. This difference in framing may influence international students' behavior when they encounter misinformation across ecologies. Moreover, the values international students have developed in their home country's digital ecologies may impact their misinformation encounters in the host country's digital ecologies [6].

Methods

Recruitment

A total of 11 students from 3 Australian universities and 5 cultural backgrounds took part in the study.

Participants were recruited via various social and student networks. Although there was the possibility of bias, recruitment posts were public. Participants were limited to international students studying in Australia to ensure that the local context was the same for all participants. Participants were required to use 2 or more social media platforms such as Facebook, Twitter, or Weibo. However, the use of platforms and immigration status were defined by the participants. Participation was voluntary, and participants were free to withdraw at any point. Participant demographics are described in Table 1.

Table 1. Demographic information of research participants (N=11).

Demographics	Values, n (%)[a]
Gender	
Female	6 (55)
Male	5 (45)
Level of study	
Master's degree	7 (64)
Bachelor's degree	2 (18)
PhD	2 (18)
Discipline	
Information systems	5 (45)
Engineering	2 (18)
Commerce	2 (18)
Medicine	1 (9)
Public health	1 (9)
Social media platforms used	
WhatsApp	11 (100)
Facebook	11 (100)
Instagram	9 (82)
LinkedIn	7 (64)
WeChat	4 (36)
Twitter	3 (27)
Weibo	3 (27)
Snapchat	2 (18)
Reddit	2 (18)
QQ	2 (18)
Line	2 (18)
WeBlock	1 (9)
Zalo	1 (9)
Country of origin	
India	4 (36)
China	3 (27)
Indonesia	2 (18)
France	1 (9)
Vietnam	1 (9)

[a]Percentages have been rounded up.

Data Collection

Data were collected in August 2021 as part of a larger study and through 3 instruments: a preinterview questionnaire, a semistructured interview, and a health misinformation-identification exercise. This research design adopts a combination of epistemological perspectives. By distinguishing the posts as information and misinformation, there is a general "truth" identified, with a post being objectively valuable because it is "true" [22]. However, especially because marginalized communities often have different experiences from those of dominant communities, truth can also be culturally constructed [23]. Interview questions were designed to elucidate how participants constructed meaning from (mis)information they came across. The objective was to understand how international students' different backgrounds and their different cultural contexts or preexisting presumptions impacted the way they interpreted information and misinformation.

The interviews were conducted and recorded using Zoom. Each of the interviews lasted 53-87 minutes. This time duration included the health misinformation-identification exercise. Interviews were conducted until 11 were complete, when data saturation was achieved. At that point, no new information emerged from the interviews. No incentives were offered to the participants.

Preinterview Questionnaire

Through the preinterview questionnaire, demographic data were collected to contextualize participants' responses during the interview. Results of this questionnaire are presented in Table 1.

Semistructured Interviews

Existing research has relied on interviews to gather data about users' experience with information [9,11]. The study used a semistructured interview format, as they are the most data dense [24]. As part of a semistructured format, participants were asked preprepared interview questions and additional follow-up questions that arose. Students were prompted to consider the questions in the context of the COVID-19 pandemic. The questions are as follows:

- Where have you encountered the same information across multiple platforms before?
- What has been your experience with the way things have been reported?
- Tell me about the last time you saw a post on a social media platform that you suspected to be misleading.

The semistructured interview explored international students' perceptions of their encounters with misleading information and the influences of inhabiting multiple digital ecologies on their consequent behaviors.

Health Misinformation-Identification Exercise

Previous literature has relied on misinformation-identification exercises [25,26]. This approach offered insights into how international students might actually react when encountering health misinformation.

Two social media posts from Australian sources were shown to the participants, one containing information and one containing misinformation about COVID-19. Claims were verified by searching for additional sources such as news articles and peer-reviewed journals to complement the original information sources. Reliability of both original and additional sources was cross-checked on Media Bias Fact Check [27].

Participants were not able to access additional sources to verify information. Participants were asked the following preprepared questions and occasional subsequent questions:

- Do you believe the information in this post is real or fake?
- How did you come to this conclusion?
- Would you want to share this information with other people?
- Why or why not?

Ethics Approval

This study was approved by The University of Melbourne's Office of Research Ethics and Integrity (2021-22022-18386-2).

Data Analysis

Similar to previous research [28,29], qualitative thematic analysis was done on the data collected between September and October 2021. Coding of the data was done by the first author (RB) in conjunction with 2 other researchers. The first author is an international student who arrived in Australia the year the pandemic started. This author was thus already sensitized to some of the experiences reported by the participants and well placed to understand the context around the participants. However, the questions asked were open-ended and designed to get participants to share their experiences rather than reflect any assumptions the first author might have had.

The coding process was initiated by the first author by reading the transcriptions as well as the interview notes several times to become familiar with the data [30]. Subsequently, an initial round of coding was done by the first author by "asking questions about the data, making comparisons between data…and in doing so, deriving concepts to stand for those data, then developing those concepts in terms of their properties and dimensions" [24]. Similar codes were then grouped together under common themes [31]. Following rounds of refinement occurred in conversation with 2 other researchers to help mitigate bias. Analogous codes were merged to distill key ideas, and themes were finessed to allow for the best possible fit of the codes and themes with the data [32]. Transcripts of interviews and evaluation exercises were reviewed again to find examples for each code derived. The codes and examples were subsequently verified by other authors to ensure there was consensus regarding the fit of the codes with the data.

Results

Overview

Analysis of the data revealed 2 key themes and 6 subthemes surrounding international students' interaction with misinformation across digital ecologies. These themes came from the concurrent analysis of data from the interviews and the exercise. Table 2 gives an overview of the data-derived themes presented in this paper. This list is not exhaustive but focuses on international students' interactions with health misinformation across digital ecologies. Table 3 provides results of the misinformation-identification exercise.

Table 2. Results from interviews with international students (N=11).

Themes and subthemes	Values, n (%)[a]
Approaches to health misinformation	
Ignoring health misinformation	8 (73)
Challenging health misinformation	5 (45)
Relying on trusted sources	4 (36)
Health misinformation across multiple digital ecologies	
Share health information with specific people	8 (73)
Evaluate health-centric posts using personal cultural values to determine reliability	7 (64)
Language barriers may inhibit uptake of host country's sources of information	3 (27)

[a]Percentages have been rounded up.

Table 3. Results from the health misinformation-identification exercise (N=11).

The exercise	Participant responses to each post, n (%)		
	True	False	Unsure
Post 1 (information)	6 (55)	3 (27)	2 (18)
Post 2 (misinformation)	0 (0)	10 (91)	1 (9)

Theme 1: Approaches to Health Misinformation

The actions international students take when they encounter health misinformation are collectively coded as "approaches to health misinformation." This includes ignoring it, challenging it, and relying on trusted sources.

Ignoring Health Misinformation

The majority of international students described at least one instance of ignoring health misinformation. They described ignoring health misinformation by choosing not to open the source, not to read it, or simply scrolling past it. P1 said, "I think my dad…was saying something like [o]h yeah, they want to track us with COVID.…I don't really react."

Challenging Health Misinformation

A little under half of the participants mentioned challenging health misinformation (5/11, 45%). Types of challenges included reporting posts, commenting on posts to share "accurate" information, or engaging the poster or sharer of misinformation in conversation. P3 said: "Most of the time, I ignore it but if it's something that relates to his health or my family's safety…[t]hen I will speak up about it."

Relying on Trusted Sources

Over a third of international students stated relying on trusted sources (4/11, 36%). Trusted sources included friends and elected officials. P8 mentioned relying on their housemates for information about COVID-19 exposure sites, saying, "I didn't bother to check the news.…I just trusted my housemates."

Theme 2: Health Misinformation Across Multiple Digital Ecologies

Health misinformation across multiple digital ecologies refers to international students' interaction with misinformation across digital ecologies grounded by different geographic locations. Subthemes show that international students share information with specific people, evaluate health-centric posts based on personal cultural values, and experience language barriers.

Share Health Information With Specific People

Most participants share health information with specific people, especially if their contacts might be affected by that information. P9 mentioned sharing information from accounts they trusted during the COVID-19 outbreak and said the following:

> *At that time, I notice right away that COVID-19 will…go with…misinformation and then you receive information from some…prestige institution. So, it would be nice.*

Evaluate Health-Centric Posts Using Personal Cultural Values to Determine Reliability

About two-thirds of international students evaluate health-centric posts using personal cultural values to determine reliability. They rely on information such as their biases against certain groups and their lived experiences. When evaluating the post in the exercise that contained COVID-19 misinformation, P2 stated, "It seems like something that insurance people back home would say."

Language Barriers May Inhibit Uptake of Host Country Social Media

For some participants, language barriers may inhibit uptake of host country social media to gather information if they do not share a first language with locals in the host country. P4 stated, "Because…I didn't check 7News like the news in Australia channel.…I just checked those Chinese version [*sic*], they would cover everything." Further, P4 explained the following:

> *I think like even though I could speak English, I read English, but I need to translate it in my brain so when I read those news, I just want to know something so*

> *I don't want to spend too much time or too much energy on that, so it's more convenient and it's much easier.*

Findings From the Health Misinformation-Identification Exercise

Findings from the exercise indicate that international students are generally reliable in identifying health misinformation. Of 11 students, 6 correctly identified post 1 as true, and 10 clearly identified post 2 as false. In both instances, participants drew on their experiences from their home countries to justify why they felt the post was true or false.

P10, when evaluating the first post, commented on the photo used in it, saying, "Yeah, because I saw AstraZeneca and I saw Covishield.…Covishield is the Indian version of AstraZeneca and AstraZeneca is from UK."

P4, when evaluating the post in the exercise that contained COVID-19 information, stated, "I don't believe the US government. They are bullshit about everything."

Generally, students who were unsure or felt that the information in post 1 was false felt so because they drew on their feelings and cultural experiences. The one student who was unsure about post 2 was concerned, as they used the same insurance referred to in the post.

Students who correctly identified the posts to be true or false drew on personal values to evaluate the content of the posts but also evaluated factors such as formatting of the posts and accounts of the posters.

Discussion

Results of this study indicate that international students navigate health misinformation across digital ecologies, and this adds complexity to their experience with health misinformation. Overall, the findings from the study complement and extend our understanding of existing research on international students' health information–seeking behaviors.

International Students Traversing Health Misinformation Across Multiple Digital Ecologies

This research affirms international students' use of multiple digital ecologies for information [4,5]. Drawing on multiple digital ecologies for information is not without challenges, and the literature highlights that language barriers may hinder international students' health information seeking in new physical environments [10,11]. This research extends our understanding by showing that language barriers also apply to their health information seeking on social media. International students' desire for information in a language they can easily interpret pushes them to sources like WeChat, where they can get tailored information from people similar to them [33,34], instead of relying primarily on their host country's sources of information.

This research shows that similar to the general population [35], international students rely on a lifetime of experience and conditioning to evaluate health information they encounter. However, for international students, this conditioning happens through use of sources in the home country digital ecologies and may then be used to make decisions about health information they encounter in digital ecologies that are not native to them.

In evaluating health information, unlike other students [36], international students evaluate the source of the health information—they rely on trusted contacts. Sometimes these contacts may be personal connections, and other times they may be official sources. In the health misinformation-identification exercise, participants evaluated the social media account sharing the information by checking whether the account was verified and checking the name of the account or news outlet to determine the credibility of the post.

However, international students do not trust all contacts. Several international students reported receiving voluminous amounts of misinformation, particularly regarding COVID-19, from home-based digital sources. In response, they ignored misinformation from their home country digital sources. Although crowd correction is vital to defending against health misinformation [37], previous literature shows that people sometimes self-censor while using social media to avoid disrupting relationships [38]. Literature also shows that if the person perceives information to be biased and influential, they may challenge it to avoid repercussions [39]. By the same token, the findings show that international students challenge health misinformation if they believe there will be adverse consequences for their family's well-being.

The health misinformation-identification exercise shows that international students can generally identify health misinformation accurately. International students' exposure to misinformation via multiple digital ecologies affects how they evaluate and react to misinformation. The COVID-19 pandemic has not only led to an increase in misinformation on social media but has also adversely impacted international students worldwide [2]. In Australia, international students were prompted to return to their home countries by the Australian government and excluded from "all federal pandemic assistance programs, such as JobSeeker and JobKeeper" [40]. The increased adversity could have resulted in worse health outcomes for international students who draw on sources from other nations. Although the general Australian population was less likely to believe COVID-19 misinformation [41], there were reports of members of multicultural and linguistically diverse communities being susceptible to misinformation about vaccines and missing out on vital health information [42,43]. However, this study has demonstrated that students who have made the digital transition to a host country ecology compare information found in their multiple ecologies. This comparison supports them in identifying health misinformation.

Multicultural communities do not all seek out information in a similar manner, and painting their experiences with one brush is harmful. International students' use of multiple ecologies may offer protection against the most negative effects of health misinformation. International students may be uniquely placed to identify it. Having exposure to multiple digital ecologies may do more than support connections with loved ones elsewhere—it may also ward against the biases present in a single ecology.

This possibility has significant implications for our understanding of how to prevent the spread of health misinformation, not just among international students but also among other groups, and it is therefore worthy of future research.

Implications

This paper sought to understand the nuances surrounding international students' interaction with health misinformation. The pandemic created an unusual impetus for international students to engage with local sources of information for up-to-date news on an evolving situation.

Findings provide insight into international students' health misinformation behaviors across digital ecologies. Further, findings reinforce that some international students face challenges in adopting host country sources of information. However, most importantly, findings combat the narrative of migrants' susceptibility to health misinformation [3,42] by showing that for international students that draw on multiple digital ecologies, there may be a protective effect against misinformation.

In consideration of the findings, health information providers need to explore ways of collaborating with digital platforms to provide vital information in a manner that is in line with how international students access health information. In addition, health information providers might also collaborate with non–health-related organizations such as community organizations and home country organizations to provide information that is easily comprehensible by those who do not share a first language with host country locals. Moreover, health organizations and educational institutes might also incorporate digital transition training into orientation sessions to help familiarize newly arrived international students with accessing health care systems and information in the host country.

Conflicts of Interest

None declared.

References

1. Gomes C, Yeoh B. Transnational Migrations in the Asia-Pacific: Transformative Experiences in the Age of Digital Media (Media, Culture and Communication in Asia-Pacific Societies). New York: Rowman & Littlefield Publishers; 2018.
2. Nguyen OTK, Balakrishnan VD. International students in Australia–during and after COVID-19. High Educ Res Dev 2020 Oct 02;39(7):1372-1376. [doi: 10.1080/07294360.2020.1825346]
3. Knights F, Carter J, Deal A, Crawshaw AF, Hayward SE, Jones L, et al. Impact of COVID-19 on migrants' access to primary care and implications for vaccine roll-out: a national qualitative study. Br J Gen Pract 2021 Apr 16;71(709):e583-e595. [doi: 10.3399/bjgp.2021.0028]
4. Gomes C, Berry M, Alzougool B, Chang S. Home away from home: international students and their identity-based social networks in Australia. JIS 2014 Jan 01;4(1):2-15. [doi: 10.32674/jis.v4i1.493]
5. Chang S, Gomes C. Digital journeys: a perspective on understanding the digital experiences of international students. JIS 2017 Apr 01;7(2):347-466. [doi: 10.32674/jis.v7i2.385]
6. Bahl R, Chang S, McKay D. Understanding international students' misinformation behavior. In: ACIS 2021 Proceedings. 2021 Presented at: ACIS; 2-5 June; Virtual conference URL: https://aisel.aisnet.org/acis2021/56
7. Marwick A. Why do People Share Fake News? A Sociotechnical Model of Media Effects. GLTR 2018;2(2):474-512 [FREE Full text]
8. Talwar S, Dhir A, Kaur P, Zafar N, Alrasheedy M. Why do people share fake news? Associations between the dark side of social media use and fake news sharing behavior. J Retail Consum Serv 2019 Nov;51:72-82. [doi: 10.1016/j.jretconser.2019.05.026]
9. Wineburg S, McGrew S. Lateral reading and the nature of expertise: reading less and learning more when evaluating digital information. Teachers College Record 2019 Nov 01;121(11):1-40. [doi: 10.1177/016146811912101102]
10. Yoon J, Kim S. Internet use by international graduate students in the USA seeking health information. Aslib J Inf Manag 2014;66(2):133. [doi: 10.1108/ajim-01-2013-0005]
11. Jang SH. "I can wait until I return home": healthcare utilization among Korean international students in the U.S. J Am Coll Health 2021 Jul 22:1-11. [doi: 10.1080/07448481.2021.1947833] [Medline: 34292859]
12. Chang S, McKay D, Caidi N, Mendoza A, Gomes C, Ekmekcioglu C. From way across the sea: information overload and international students during the COVID-19 pandemic. Proc Assoc Inf Sci Technol 2020 Oct 22;57(1):e289 [FREE Full text] [doi: 10.1002/pra2.289] [Medline: 33173816]
13. Wu L, Morstatter F, Carley KM, Liu H. Misinformation in social media. SIGKDD Explor Newsl 2019 Nov 26;21(2):80-90. [doi: 10.1145/3373464.3373475]
14. Boyd D, Ellison N. Social network sites: definition, history, and scholarship. JCMC 2007;13(1):210-230. [doi: 10.1111/j.1083-6101.2007.00393.x]
15. Ortega G, Rodriguez JA, Maurer LR, Witt EE, Perez N, Reich A, et al. Telemedicine, COVID-19, and disparities: policy implications. Health Policy Technol 2020 Sep;9(3):368-371 [FREE Full text] [doi: 10.1016/j.hlpt.2020.08.001] [Medline: 32837888]

16. Sawir E, Marginson S, Deumert A, Nyland C, Ramia G. Loneliness and international students: an Australian study. J Stud Int Educ 2007 Sep 13;12(2):148-180. [doi: 10.1177/1028315307299699]
17. Erdelez S. Information encountering: it's more than just bumping into information. Bull Am Soc Inform Sci Technol 2005 Jan 31;25(3):26-29. [doi: 10.1002/bult.118]
18. Fletcher R, Nielsen RK. Are people incidentally exposed to news on social media? A comparative analysis. New Media Soc 2017 Aug 17;20(7):2450-2468. [doi: 10.1177/1461444817724170]
19. 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]
20. Pennycook G, Cannon TD, Rand DG. Prior exposure increases perceived accuracy of fake news. J Exp Psychol Gen 2018 Dec;147(12):1865-1880. [doi: 10.1037/xge0000465]
21. Porshnev A, Miltsov A, Lokot T, Koltsova O. Effects of Conspiracy Thinking Style, FramingPolitical Interest on Accuracy of Fake News Recognition by Social Media Usersvidence from Russia, KazakhstanUkraine. In: Social Computing and Social Media: Experience Design and Social Network Analysis. 2021 Presented at: 13th International Conference, SCSM 2021, Held as Part of the 23rd HCI International Conference, HCII 2021; 24-29 July; Virtual Conference. [doi: 10.1007/978-3-030-77626-8_23]
22. Bernecker S, Flowerree A, Grundmann T. The Epistemology of Fake News. Oxford: Oxford University Press; 2021.
23. Ruokolainen H, Widén G. Conceptualising misinformation in the context of asylum seekers. Inf Process Manage 2020 May;57(3):102127. [doi: 10.1016/j.ipm.2019.102127]
24. Corbin J, Strauss A. Basics of Qualitative Research (3rd ed.): Techniques and Procedures for Developing Grounded Theory Internet. California, US: SAGE Publications; 2008.
25. Leeder C. How college students evaluate and share “fake news” stories. Libr Inf Sci Res 2019 Jul;41(3):100967. [doi: 10.1016/j.lisr.2019.100967]
26. Tarchi C. Identifying fake news through trustworthiness judgements of documents / La identificación de noticias falsas mediante juicios de fiabilidad de los documentos. Cultura y Educación 2019 May 10;31(2):369-406. [doi: 10.1080/11356405.2019.1597442]
27. Media Bias Fact Check. URL: https://mediabiasfactcheck.com/ [accessed 2022-06-21]
28. Fletcher R, Nielsen RK. Generalised scepticism: how people navigate news on social media. Inf Commun Soc 2018 Mar 20;22(12):1751-1769. [doi: 10.1080/1369118x.2018.1450887]
29. Binsahl H, Chang S, Bosua R. Cross-cultural digital information-seeking experiences. JIS 2020 Nov 15;10(4):872-891. [doi: 10.32674/jis.v10i3.1573]
30. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
31. Kiger ME, Varpio L. Thematic analysis of qualitative data: AMEE Guide No. 131. Med Teach 2020 Aug 01;42(8):846-854. [doi: 10.1080/0142159X.2020.1755030] [Medline: 32356468]
32. Braun V, Clarke V. Successful Qualitative Research: A Practical Guide for Beginners. California, US: Sage Publications; 2013.
33. Agosto DE, Hughes-Hassell S. People, places, and questions: an investigation of the everyday life information-seeking behaviors of urban young adults. Libr Inf Sci Res 2005 Mar;27(2):141-163. [doi: 10.1016/j.lisr.2005.01.002]
34. Chang S, Gomes C, Platt M, Trumpour S, McKay D, Alzougool B. Mapping the contours of digital journeys: a study of international students’ social networks in Australian higher education. High Educ Res Dev 2021 Aug 14:1-17. [doi: 10.1080/07294360.2021.1962812]
35. Currie Sivek S. Both facts and feelings: emotion and news literacy. JMLE 2018;10(2):123-138. [doi: 10.23860/jmle-2018-10-2-7]
36. Hansen DL, Derry HA, Resnick PJ, Richardson CR. Adolescents searching for health information on the Internet: an observational study. J Med Internet Res 2003 Oct 17;5(4):e25 [FREE Full text] [doi: 10.2196/jmir.5.4.e25] [Medline: 14713653]
37. Bode L, Vraga E. The Swiss cheese model for mitigating online misinformation. Bull At Sci 2021 May 11;77(3):129-133. [doi: 10.1080/00963402.2021.1912170]
38. Seargeant P, Tagg C. Social media and the future of open debate: a user-oriented approach to Facebook’s filter bubble conundrum. Discourse Context Me 2019 Mar;27:41-48. [doi: 10.1016/j.dcm.2018.03.005]
39. Koo AZ, Su M, Lee S, Ahn S, Rojas H. What motivates people to correct misinformation? Examining the effects of third-person perceptions and perceived norms. J Broadcast Electron Media 2021 Apr 20;65(1):111-134. [doi: 10.1080/08838151.2021.1903896]
40. Weng E, Halafoff A, Barton G, Smith G. Higher education, exclusion, and belonging. JIS 2021 Sep 09;11(S2). [doi: 10.32674/jis.v11iS2.3553]
41. Pickles K, Cvejic E, Nickel B, Copp T, Bonner C, Leask J, et al. COVID-19 misinformation trends in Australia: prospective longitudinal national survey. J Med Internet Res 2021 Jan 07;23(1):e23805 [FREE Full text] [doi: 10.2196/23805] [Medline: 33302250]

42. Yussuf A. Anti-vax 'fearmongers' spreading misinformation are targeting Australia's diverse communities, leaders, experts warn. ABC News. 2021 Sep 5. URL: https://www.abc.net.au/news/2021-09-06/fears-misinformation-targeting-australias-diverse-communities-/100405706 [accessed 2022-06-21]
43. Grey A. Australia's multilingual communities are missing out on vital coronavirus information. ABC News. 2020 Jun 29. URL: https://www.abc.net.au/news/2020-06-29/coronavirus-multilingual-australia-missing-out-covid-19-info/12403510 [accessed 2022-06-21]

Edited by M Gisondi, M Strehlow; submitted 05.04.22; peer-reviewed by PCI Pang, SF Tsao; comments to author 03.05.22; revised version received 24.05.22; accepted 31.05.22; published 06.07.22.

Please cite as:
Bahl R, Chang S, McKay D, Buchanan G
Health Misinformation Across Multiple Digital Ecologies: Qualitative Study of Data From Interviews With International Students
J Med Internet Res 2022;24(7):e38523
URL: https://www.jmir.org/2022/7/e38523
doi:10.2196/38523
PMID:35727960

Original Paper

Exploring COVID-19–Related Stressors: Topic Modeling Study

Yue Tong Leung[1], MEng; Farzad Khalvati[1,2,3,4], PhD

[1]Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, ON, Canada
[2]Department of Diagnostic Imaging, The Hospital for Sick Children, Toronto, ON, Canada
[3]Department of Medical Imaging, University of Toronto, Toronto, ON, Canada
[4]Institute of Medical Science, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Farzad Khalvati, PhD
Department of Diagnostic Imaging
The Hospital for Sick Children
555 University Avenue
Toronto, ON, M5G 1X8
Canada
Phone: 1 416 813 7910
Email: Farzad.Khalvati@utoronto.ca

Abstract

Background: The COVID-19 pandemic has affected the lives of people globally for over 2 years. Changes in lifestyles due to the pandemic may cause psychosocial stressors for individuals and could lead to mental health problems. To provide high-quality mental health support, health care organizations need to identify COVID-19–specific stressors and monitor the trends in the prevalence of those stressors.

Objective: This study aims to apply natural language processing (NLP) techniques to social media data to identify the psychosocial stressors during the COVID-19 pandemic and to analyze the trend in the prevalence of these stressors at different stages of the pandemic.

Methods: We obtained a data set of 9266 Reddit posts from the subreddit \rCOVID19_support, from February 14, 2020, to July 19, 2021. We used the latent Dirichlet allocation (LDA) topic model to identify the topics that were mentioned on the subreddit and analyzed the trends in the prevalence of the topics. Lexicons were created for each of the topics and were used to identify the topics of each post. The prevalences of topics identified by the LDA and lexicon approaches were compared.

Results: The LDA model identified 6 topics from the data set: (1) “fear of coronavirus,” (2) “problems related to social relationships,” (3) “mental health symptoms,” (4) “family problems,” (5) “educational and occupational problems,” and (6) “uncertainty on the development of pandemic.” According to the results, there was a significant decline in the number of posts about the “fear of coronavirus” after vaccine distribution started. This suggests that the distribution of vaccines may have reduced the perceived risks of coronavirus. The prevalence of discussions on the uncertainty about the pandemic did not decline with the increase in the vaccinated population. In April 2021, when the Delta variant became prevalent in the United States, there was a significant increase in the number of posts about the uncertainty of pandemic development but no obvious effects on the topic of fear of the coronavirus.

Conclusions: We created a dashboard to visualize the trend in the prevalence of topics about COVID-19–related stressors being discussed on a social media platform (Reddit). Our results provide insights into the prevalence of pandemic-related stressors during different stages of the COVID-19 pandemic. The NLP techniques leveraged in this study could also be applied to analyze event-specific stressors in the future.

(J Med Internet Res 2022;24(7):e37142) doi:10.2196/37142

KEYWORDS

COVID-19; natural language processing; public health informatics; topic modeling

Introduction

The COVID-19 pandemic has affected the lives of people globally for over 2 years. To mitigate infection, safety measures such as social distancing, lockdowns, and school closures have been implemented. The mental health of many has been impacted due to stress, anxiety, loneliness, and feelings of uncertainty about the pandemic [1]. Recognizing the common psychosocial stressors and their prevalence at different stages of the pandemic will allow health care providers and social workers to provide better mental health interventions and support. Several studies have been conducted to understand the COVID-19–related stressors and the impacts of COVID-19 on mental health [1-3]. The majority of the studies used questionnaires or interviews to obtain data. However, the process of obtaining data could be time-consuming and costly. As the pandemic is evolving rapidly, the prevalence of stressors may also be changing due to different factors such as the social distancing measures, coronavirus infection rates, and vaccine distribution. Survey methods may not be able to capture current COVID-19–related stress. In contrast, social media may allow close to real-time monitoring of the changes in stressors during the pandemic, as individuals actively share their feelings and difficulties on social media platforms. In this study, natural language processing (NLP) techniques are applied to identify COVID-19–related psychosocial stressors that are discussed on social media and to visualize the prevalence of stressors at different stages of the pandemic.

Social media platforms, such as Twitter and Reddit, are commonly used as the data source for obtaining insights regarding mental health status. As people share their feelings or experiences on the platforms, the content may reflect users' emotions. The changes in emotions of the population could be reflected by their behavior on social media. Many researchers have utilized NLP techniques and social media data to analyze the mental health status of the population. De Choudhury et al [2] introduced a social media depression index to monitor the trends in depression across the population. Larsen et al [3] devised a system, called "We Feel," that applies sentiment analysis and data visualization to tweets to obtain real-time insights into the emotional state of the population.

In the field of public health and informatics, some researchers have utilized NLP topic modeling to summarize COVID-19–related discussions on social media. Medford et al [4] observed the trend in topics in COVID-19–related tweets in the early stage of the outbreak. Jelodar et al [5] leveraged topic modeling to identify the topics that were being discussed on COVID-19–related subreddits such as \rCOVID19, \rCoronavirus, and r/CoronaVirus2019nCoV. Jang et al [6] used topic modeling to identify topics on Twitter to understand the reactions and concerns about COVID-19 in North America. Besides summarizing pandemic-related discussions, some research has leveraged NLP techniques on social media data to analyze the mental health impacts of COVID-19. Biester et al [7] observed changes in Reddit mental health support groups after the COVID-19 outbreak and obtained insights into the impact on mental health. In that study, the topic model was applied to identify topics such as family and school in the subreddits; then, time series analysis was leveraged to find which of the topics were affected after the outbreak of the pandemic. Low et al [8] used NLP techniques to characterize the differences in Reddit mental health support groups in the prepandemic and mid-pandemic periods.

Although some prior NLP research aimed at devising methods to utilize social media data to measure the population's mental health status, few studies focused on identifying psychosocial stressors. According to the American Psychological Association [9], a psychosocial stressor is defined as "a life situation that creates an unusual or intense level of stress that may contribute to the development or aggravation of mental disorder, illness, or maladaptive behavior." Mowery et al [10] developed an annotation scheme to identify depression symptoms and psychosocial stressors, such as problems with expected life course and problems with the primary support group, mentioned in tweets. Regarding the stressors during the pandemic, existing research used traditional methods such as questionnaires. Park et al [11] developed a questionnaire-based assessment tool for COVID-19–related stressors. The questions included whether individuals experienced the following in the past week: "risk of becoming infected," "risk of unintentionally infecting other people," "cancellation of meaningful personal or religious rituals," and "loss of current job security or income." People experiencing those stressors may also share their experiences on social media. If we could identify the stressors that were mentioned in posts, we would have real-time monitoring of changes. Our study aimed to provide an alternative method using NLP and social media data to obtain related data and summarize the prevalence of COVID-19–related stressors.

In this paper, we utilized the latent Dirichlet allocation (LDA) topic model to identify pandemic-related distress, by identifying the topics being discussed on the subreddit \rCOVID19_support. After applying the LDA model, we visualized the trends in the prevalence of topics at different stages of the pandemic. Several existing works leveraged NLP to analyze the population's mental health status during the pandemic; however, existing studies did not explore the possibility of analyzing COVID-19–related stressors. This study focused on monitoring the stressors during the pandemic. Although most existing research focused on the mental health impacts at the beginning of the outbreak of the pandemic, the data set extracted in this study covered the posts on the subreddit starting from February 2020, the outbreak of the pandemic, up to July 2021. This allowed us to visualize the changes in the prevalence of stressors at different stages of the pandemic. Observing the trends can provide insights into the latest predominant stressors, and the findings could also be useful for mental health support providers and policy makers. We believe applying NLP to summarize text can help us obtain insights into mental health status and stressors during the pandemic.

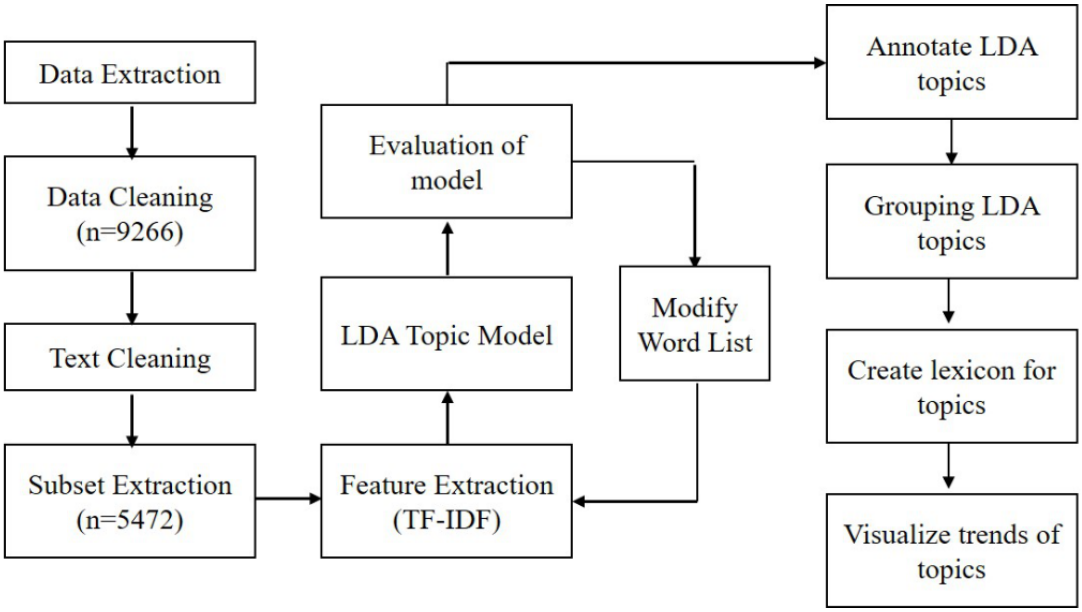
Methods

Study Design

Figure 1 describes the overview of the study. First, we extracted the Reddit posts on \rCOVID19_support and preprocessed the data, including data cleaning and text cleaning. We then

extracted a subset of posts that were likely to include content related to our research topic. Next, we created the LDA topic model to cluster the posts into different LDA topics for identifying the topics that existed in our data set and grouped the LDA topics with similar content into topic groups. By reviewing the keywords and example posts in those topic groups, we gained insights into the popular topics that existed in the data set. Once we identified the popular topics in our data set, we observed the frequently appearing words in each of the topics and identified the keywords for the topics. Next, we created psychosocial stressor lexicons, which were sets of keywords for each of the topics about the stressors. We then used the LDA topic model and the lexicons to label the topics that were covered in the posts. With the labels, we measured the number of posts with those topics for each month within our study period. Then, we could visualize the trends in the prevalence of topics on the subreddit.

Figure 1. Overview of the research framework. LDA: latent Dirichlet allocation; TF-IDF: term frequency-inverse document frequency.



Data Collection and Preprocessing

In this study, Reddit was selected as the data source for applying machine learning models to obtain insights into the prevalence of stressors during COVID-19. There are several advantages of using Reddit in this study. De Choudhury and De [12] found that the anonymity of throwaway accounts on forums like Reddit can promote self-disclosure about mental health issues. Naseem et al [13] found that COVID-19–related discussions on Twitter focused on news and opinions about government policies, while the discussions on Reddit focused on how the pandemic affected people's lives. As a result of these tendencies, Reddit has a high potential to be a better data source for identifying psychosocial stressors during the pandemic.

There are different subreddits related to COVID-19. We selected \rCOVID19_support for studying COVID-19-related stressors. Users on the platform \rCOVID19_support ask questions about COVID-19 and share their experience during hard times in the pandemic. The discussions on other subreddits, such as \rCOVID-19, \rCoronavirus, and \rCoronaVirus2019nCov focused on news and information instead of sharing experiences during the pandemic. As the topics discussed on \rCOVID19_support are more related to our research topic, we selected this subreddit as our data source.

Reddit posts were extracted from the subreddit \rCOVID19_support using the Pushshift API [14]. The data set included 9266 posts from February 14, 2020, to July 19, 2021. The posts marked as "[removed]" or "[deleted]" were excluded from the data set. After data cleaning, we applied the following preprocessing steps to avoid the noise influencing the topic model: First, we joined the texts in the titles and content of the posts; second, we removed the hyperlinks in the posts by removing tokens that included characters "www"; third, we replaced words spelled informally with the formal spelling, for example "ive" to "I have" and "ppl" to "people"; fourth, we lemmatized the words; fifth, we removed stop words, such as "would," "they," and "are" in the text; sixth, we removed punctuation and numbers.

On Reddit, some of the posts are tagged by "flairs," which describe the content or the nature of the posts. The flairs in the data set included: "Support," "Questions," "Discussions," "Trigger Warning," "Good News," "Firsthand Account," "Resources," "Vaccines are SAFE," "News," "Biosafety Request," "The answer is NO," "Misinformation-debunked," and "Desperate mod."

Posts that are tagged with flairs such as "Resources" and "News" have a low tendency to include content related to psychosocial stressors. With the use of flairs tagged to the posts, we filtered posts with a low chance of including content related to our research topic. On the subreddit, some of the posts included content related to sharing information such as the latest news about COVID-19, potential adverse effects of vaccines, and tips about infection prevention. Some users asked questions such as which vaccines are safe, whether the adverse effects of vaccines are normal, and whether it is safe to visit their grandparents during the pandemic. The posts labelled with flairs

"Support" and "Trigger Warning" have a high tendency to include content about users' personal experiences, stressors, or feelings during the pandemic. Posts with flairs "News" and "Questions" tend to not include content about stressors. As we were focused on understanding stressors in this study, we extracted a subset to include posts that were labelled with the flairs that have a high tendency to include relevant content.

In the data set, 4654 posts were labelled by flairs, and 4612 posts were not. The number of posts tagged by each of the flairs is shown in Table 1. Missing flairs were predicted using a logistic regression model that was trained by the labelled data.

Before training the classifier, the texts in the posts were represented by the term frequency-inverse document frequency (TF-IDF), which quantifies the significance of a given term compared with the other terms within the given document and within the given corpus [15]. The features used to train the classifier included LDA features (n=10), TF-IDF features (n=200), and 1 feature to describe whether the posts included hyperlinks (n=1). After filling in the missing flairs in the data set, 5472 data points were labelled by flairs that likely described stressors during COVID-19. These data points were extracted as a subset for analysis.

Table 1. Number of posts tagged by each of the flairs in the data set.

Flairs	Subset with labelled flairs (n=4654)	Subset with predicted flairs (n=4612)	Data set with labelled or predicted flairs (n=9266)
Mental health support			
Support	2386	—[a]	—
Trigger warning	197	—	—
Deperate mod[b]	1	—	—
Total	2584	2888	5472
Discussion and questions			
Questions	1069	—	—
Discussion	597	—	—
Vaccines are SAFE	55	—	—
The answer is NO	7	—	—
Biosafety request	14	—	—
Total	1742	1417	3159
News and resources			
Good news	146	—	—
Resources	59	—	—
News	18	—	—
Misinformation—debunked	3	—	—
Total	226	225	451
Experience			
Firsthand account	102	—	—
Total	102	82	184

[a]Not applicable.

[b]The flair in the original post was "Deperate mod," which is likely a misspelling of "desperate mood."

Topic Model Training and Evaluation

In this study, we used the LDA topic model to identify the COVID-19–related distress that was mentioned on Reddit. The topic model is an unsupervised machine learning model that can be applied to different research topics such as computational social science and understanding scientific publications [16]. LDA is a topic model that is commonly used to summarize the topics on social media. The LDA model is based on the following assumptions: A topic is a combination of terms with a probability distribution, and a document is generated by a combination of topics with a probability distribution [16]. The algorithm starts by setting topic assignments randomly and then computes the distribution of words in a topic and the distribution of topics in a document. Then, the topic allocation of words is updated iteratively until convergence. The LDA model returns the distribution of topics in each of the documents and the distribution of words in each of the topics. With the topic allocation of documents, the dominant topic of each of the posts in our data set could be determined.

In this study, the LDA model in the Python scikit-learn package was used to identify the topics in the data set. The model was trained by TF-IDF features that were created from the data set.

For each of the data points, the topic model outputs the proportions of contents belonging to each of the topics. The dominant topic for each of the posts was then identified by finding the topic with the highest value in the LDA output. The number of topics is the major hyperparameter of the LDA model, which affects the interpretability of topics. There are different ways to determine a suitable number of topics to achieve high interpretability of topics. In the study by Jelodar et al [5], the number of topics was first set to a high number then clustered into 18 topics by human judgement. In our study, we used a similar method to select the number of topics. First, we set the number of LDA topics to 100 and checked the dominant topic of each of the posts and found that only 25 LDA topics had more than 10 posts with corresponding topics. Then, we analyzed each of the 25 LDA topics and clustered them into 6 topics using human judgment. For each of the LDA topics, 3 sample posts with the highest LDA output percentage and 3 random posts with corresponding dominant topics were selected for manual review. Topic interpretability was manually evaluated by reviewing the data points selected. If the sample data points of the same dominant topics had similar contents, the topic could be named. After naming the LDA topics, some topics shared similar content with other topics. The LDA topics were then grouped into "topic groups." We identified 6 topic groups: educational and occupational problems, family problems, fear of coronavirus, mental health symptoms, problems related to social relationships, and uncertainty about the development of the pandemic. The UMass coherence score of the LDA topic model is –2.106.

Feature Engineering

The output of topic modeling is highly dependent on the feature vector (a matrix of TF-IDF values for each of the documents). For the first trial, the feature vector used to train the LDA model included TF-IDF with max_feature 300. This means the feature vector includes 300 columns of tokens, which are the terms consisting of one or more words that have the highest TF-IDF values in the given corpus. Then, the model was evaluated by the aforementioned methods, namely selecting sample posts from each of the topics and evaluating the topic coherence manually. In addition to evaluating LDA topic coherence, features were manually evaluated to determine whether they were likely to cause the LDA topic model to cluster posts in desired ways. For example, the topic model may group posts with tokens "suggestion," "anyone," and "thank you" into the same topic because those words tend to appear together when authors were asking for suggestions at the end of posts. However, clustering this topic does not help with understanding the stressors and may act as noise. To avoid this, those tokens were removed from the feature vectors. Besides removing those tokens, we also identified some tokens that were useful for identifying the topics. For example, the tokens "grocery shopping," "maskless," and "no mask" commonly appeared when users were expressing their fears of getting infected. We hypothesized that including those tokens in the feature vector would help the LDA topic model identify topics related to fear of coronavirus. The feature vector was then updated by selecting these tokens. Then, the LDA was trained by the new feature vector, leading the output of the model to be closer to the desired result. The iteration process to improve the topic model is illustrated in Figure 1. The iteration ended when the sample posts shared similar content within the same LDA topic. Using this iterative approach to feature selection and evaluation of the topic model output, the performance of the topic model was optimized.

Psychosocial Stressor Lexicon

A psychosocial stressor lexicon is a list of keywords for each of the stressors. After the topic groups were defined, each of the topic groups was evaluated to search keywords that directly indicate the existence of topics in the text. For example, if the posts include education-related words such as "college" and "online learning," it can be concluded that the authors have mentioned educational problems in the post. Lexicons were created for some of the topics. If a post included any of the words listed in the lexicons, it was assumed that the post included content related to the corresponding topics. Each of the posts could contain more than one topic. With the use of lexicons, each of the posts was annotated by whether it included content of each of the topics.

In LDA, topics are defined as a mixture of terms with different probability distributions; this means a word could belong to more than one topic and it could cause inaccuracy in the prediction. In contrast, the lexical approach has higher interpretability on topic classification, but it requires careful selection of keywords to avoid including terms belonging to more than one topic. In this study, we assumed that we do not have prior knowledge of how Reddit users expressed their feelings on the subreddit. To obtain insights into what topics existed in the subreddit and the keywords for each of the topics, we applied the LDA model before applying the lexical approach. In this study, the lexical approach was created for 2 purposes: first, to further analyze the subtopics within the topic group. Some of the topics may include some common words and have to be grouped into the same topic group in the LDA model. The topics in the same topic group could be separated by choosing unique keywords. The second reason for creating the lexicon was to verify the result from the LDA.

Visualization

In previous steps, each of the data points was annotated with the LDA output, which represents the proportion of the content belonging to each of the topic groups, and the results of the lexical approach, which represent whether the post includes words that were listed in lexicons. Then, the monthly sum of the LDA model output for each topic groups was computed. The trend of topics in the data set can then be visualized and compared with the development of the pandemic. Regarding the pandemic development, the numbers of total cases, new cases per day, and vaccinated population were obtained from Our World Data [17]. In this study, only the numbers from the United States, the United Kingdom, and Canada are included because the majority of Reddit users are from these countries [18].

Ethical Considerations

This study uses public datasets derived from social media and does not involve any human participants. As such it does not require ethics board approval.

Results

Topics Identified by the Topic Model

After grouping LDA topics, 6 topics were identified: (1) fear of coronavirus, (2) educational and occupational problems, (3) family problems, (4) problems related to social relationships, (5) mental health symptoms, and (6) uncertainty about the development of the pandemic. We grouped the posts according to the topics and created word clouds for each of the 6 subsets. Figure 2 shows the word clouds, which describe the frequently used words for each of the topics. The font size in the word clouds represents the frequency of words in the topics. By visualizing the keywords, the word clouds summarized the content for each of the topics.

We used t-distributed Stochastic Neighbor Embedding (t-SNE), which can be used to visualize high-dimensional data, to visualize the coherence of topic groups.

Figure 3 shows the t-SNE that describes the clustering of topic groups in our data. As the posts could include multiple topics, we could observe visible crossover between the topic clusters in the t-SNE. It shows the topics “fear of coronavirus” and “family problem” are similar. This could be because many posts related to family problems are about the worries of infecting family members and worries about the situations of parents who were infected.

The major topic in the data set was “fear of coronavirus.” The words frequently appearing in the topic of “fear of coronavirus” are shown in Figure 2. The words “home,” “live,” and “family” appeared frequently when the authors of posts explained they were living with family members, followed by expressing worries of infecting or getting infected by the family members. The words “mask,” “risk,” and “scare” appeared when the authors shared their worries about getting infected outside of their home.

The LDA model identified the topic of “educational and occupational problems.” Some students shared their feelings about the online learning experience, and some felt upset because they missed the social engagement from school activities. Some people expressed worries about finding jobs during the pandemic. Other posts mentioned the situation of current university students having online classes for the final year and worried about finding a job after graduation. Due to the existence of this type of content and tokens related to education and work such as “online learning,” “graduation,” “university,” and “find jobs” appearing together, we grouped the educational and occupational LDA topics into the same group.

The topic “family problems” included content about several subtopics: the worry of elderly parents getting infected, worry about parents who were already infected, and the anger of having opposing views with family members due to different acceptance of social distancing measures. The topic “mental health symptoms” included mentioning symptoms such as insomnia, obsessive compulsive disorder, and feeling depressed. Some of the users described their feelings and mental health symptoms on Reddit and asked for mental health support but did not explain the stressors. The topic “problems related to social relationships” included content about loneliness and uncertainty about the balance between social activities and pandemic safety measures. The topic “uncertainty on development of pandemic” refers to the content about the worry of the pandemic lasting forever. For this topic, some posts included discussions about whether vaccination would help get life back to normal.

Figure 2. Word clouds for each of the topics identified: (A) uncertainty about the development of the pandemic, (B) problems related to social relationships, (C) family problems, (D) educational and occupational problems, (E) mental health symptoms, (F) fear of the coronavirus.

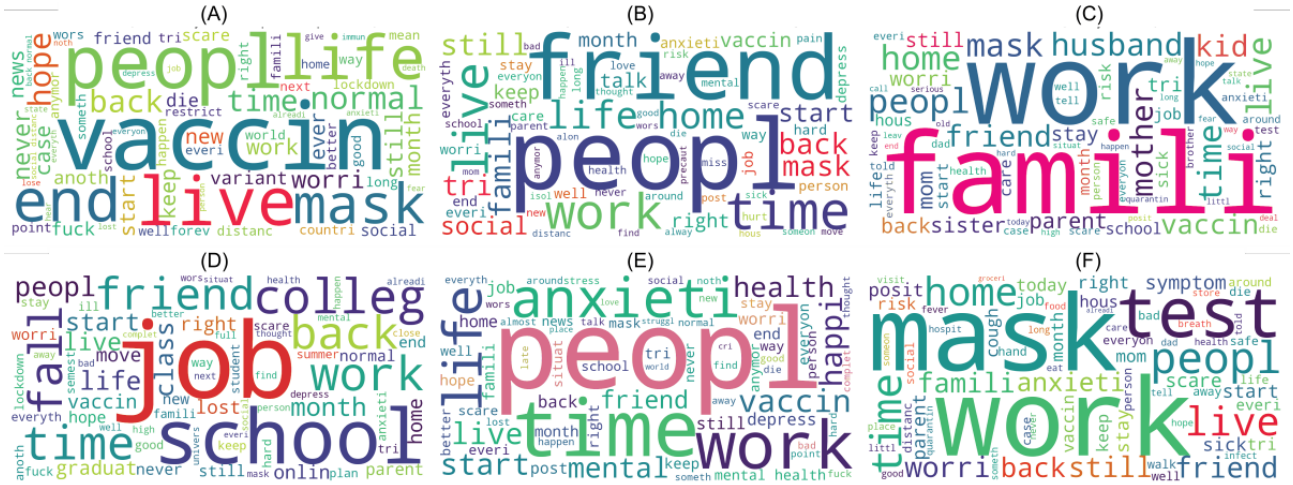
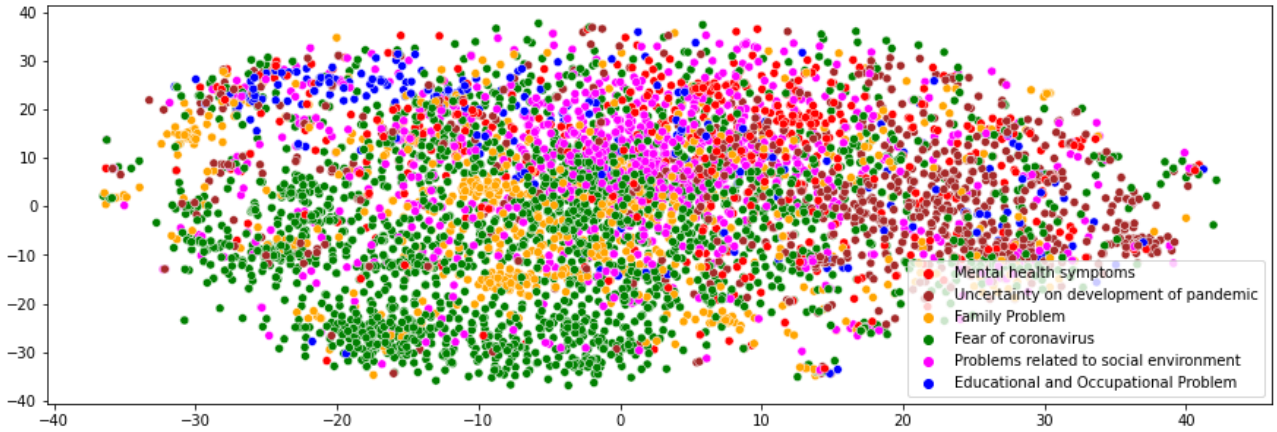


Figure 3. t-distributed Stochastic Neighbor Embedding (t-SNE) plot for the topic groups identified.



Trends in LDA Topic Groups

On March 11, 2020, the World Health Organization declared COVID-19 a "pandemic" [19]. As shown in Figure 4, the number of posts on the subreddit was the highest in March 2020. The number declined from April 2020 to June 2020, then rose steadily from September 2020 to December 2020. The number of posts related to stressors then declined in February 2021, and vaccine distribution started between January 2021 and February 2021.

The trends in the prevalence of topics have been plotted separately in Figure 5. To understand the relationship between the trends in stressors and the number of COVID-19 cases, the daily number of cases was also included.

As shown in Figure 5, the number of posts mentioning the fear of coronavirus did not significantly change from May 2020 to December 2020, although the daily number of cases had a significant increase in November 2020. This suggests the fear of coronavirus was independent of the actual number of cases. The number of posts with the topic of "fear of coronavirus" was the highest in March 2020, when COVID-19 was first declared a "pandemic." In February 2021, the discussions about the fear of coronavirus significantly reduced. This suggests the distribution of vaccines may have reduced the perceived risks of infection with the coronavirus.

The prevalence of the topics "family problems" and "educational and occupational problems" dropped in February 2021. For the topic of uncertainty, its prevalence declined in February 2021 but increased in April 2021, when the new variant was becoming prevalent in the United States. The most common topic in December 2020 and January 2021 was "uncertainty about development of pandemic." Starting from May 2020, the prevalence had a high correlation with the number of cases.

To determine the distribution of topics in each month, we calculated the percentage of posts belonging to each topic in each month, and the corresponding trend was plotted (Figure 6).

Figure 6 displays the proportion of each topic with respect to all topics. As shown, "Fear of coronavirus" was the major topic of discussion on the platform. From March 2020 to November 2020, more than 40% of the content on the forum belonged to the fear of coronavirus topic. Starting from December 2020, the proportion was below 40% and showed a decreasing trend, while the proportion of posts about "Uncertainty on development of pandemic" steadily increased. In June 2020, the proportion of the topics about pandemic development was higher than the topics about fear. For other topics, the prevalence did not significantly change. The results suggest that the major stressor related to COVID-19 shifted from the fear of getting infected to a feeling of uncertainty about the development of the pandemic.

Figure 4. Trend of topics on \rCOVID19_support. The stacked area plot represents the sum of the latent Dirichlet allocation (LDA) output for each month, for each of the topics. The line plot represents the total number of cases and the vaccinated population.

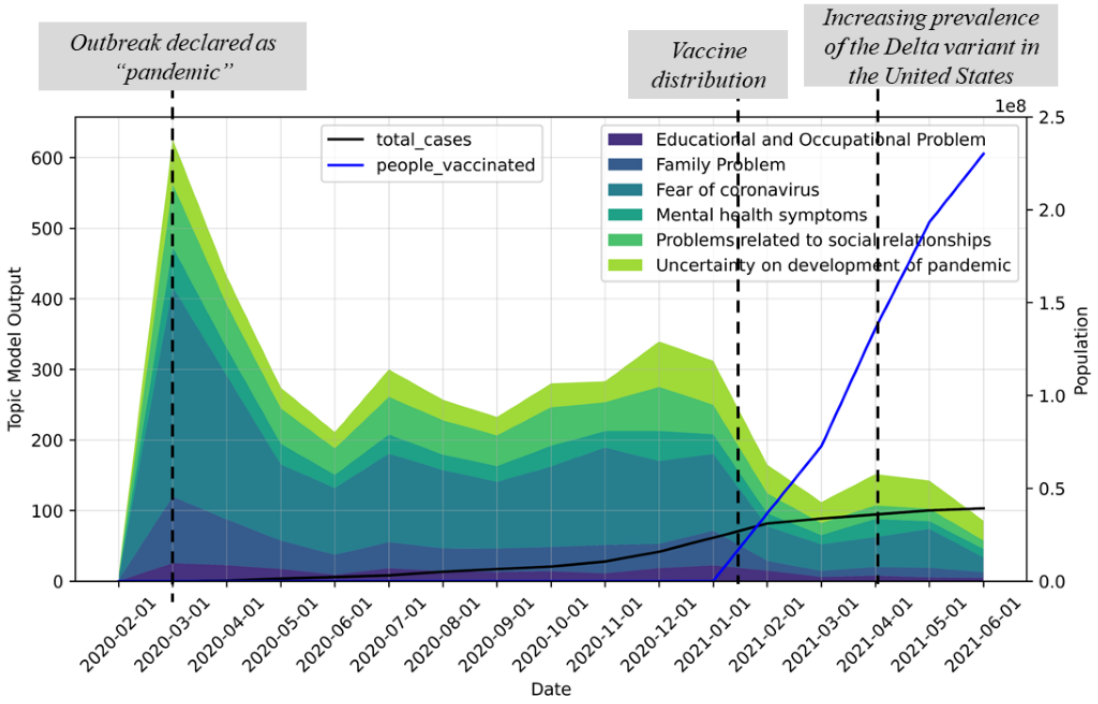


Figure 5. Trends in the prevalence of topic groups: (A) educational and occupational problems, (B) mental health symptoms, (C) family problems, (D) problems related to social relationships, (E) fear of the coronavirus, and (F) uncertainty on the development of the pandemic. The line plots represent the number of cases for each month. The area plots represent the prevalence of topics, which was measured using the output of the latent Dirichlet allocation (LDA).

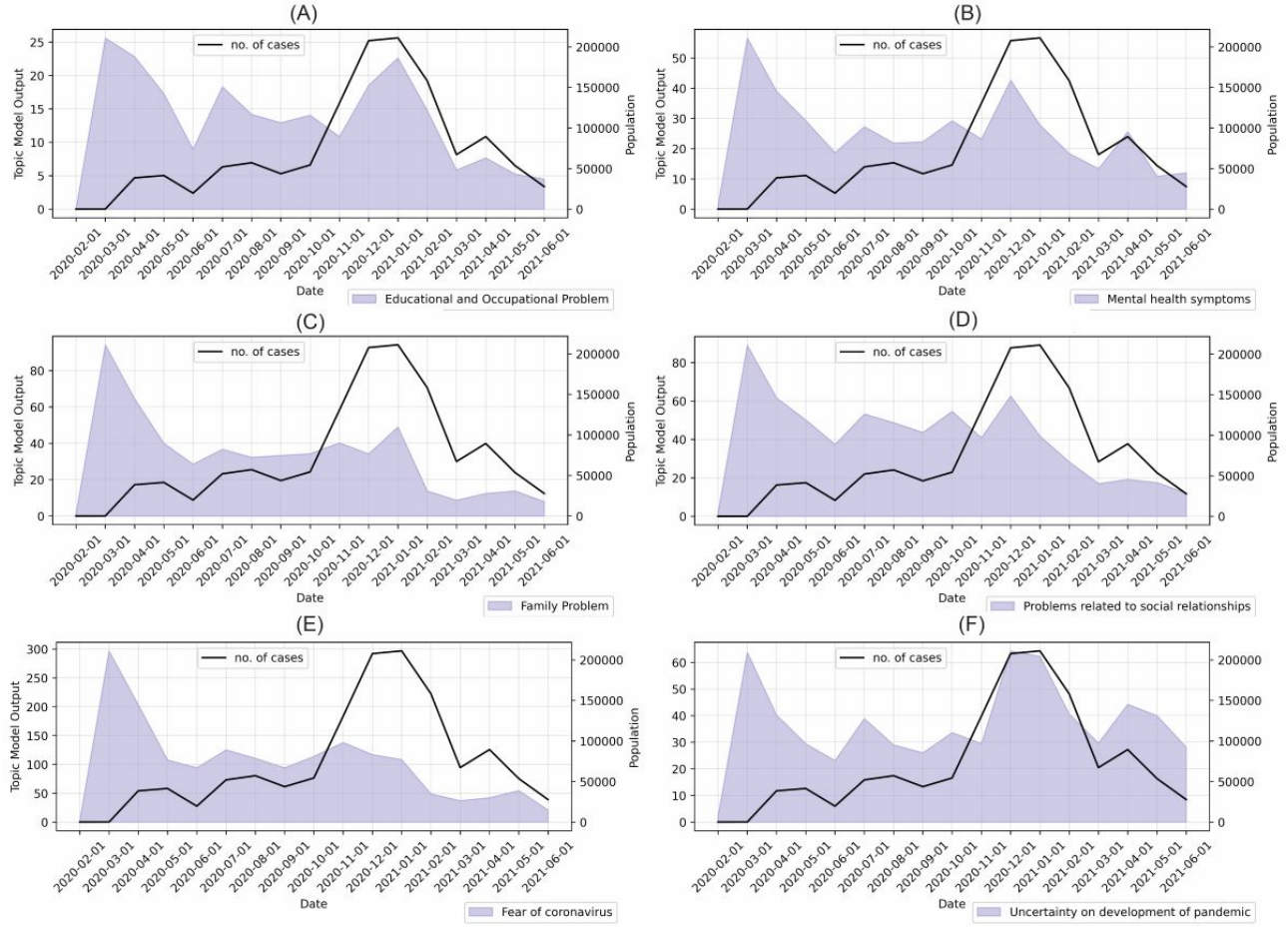
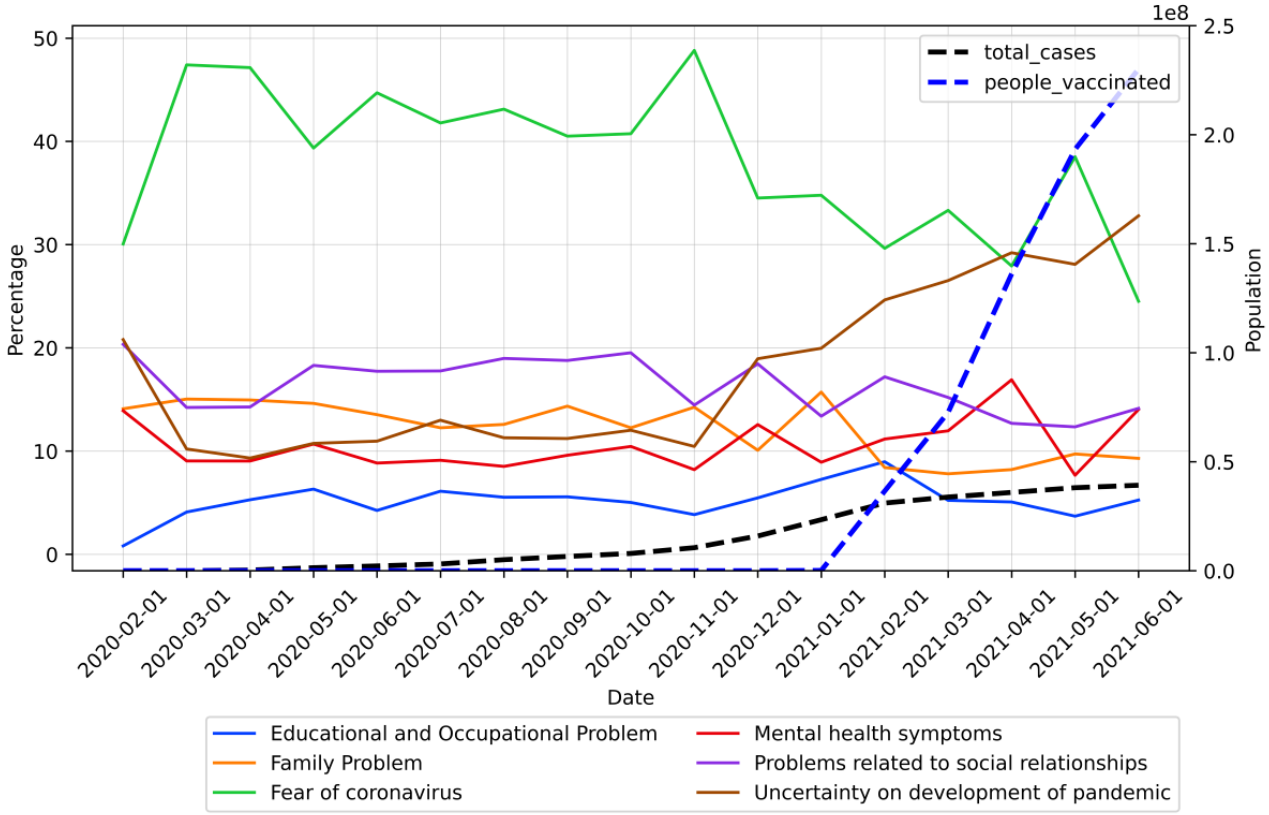


Figure 6. Trend on proportions of topics mentioned on \rCOVID19_support. The dashed lines represent the total number of cases and the vaccinated population. The solid lines represent the proportion of each topic for each month.



Lexicon Approach

Table 2 shows the lexicon created based on the result of the LDA model. The lexicons were created for 2 purposes: first, to further analyze the subtopics within the topic group. As explained in previous sections, the LDA topics of occupational problems and educational problems had to be grouped. By reviewing the word clouds and sample texts of the topic group, we listed the words that could identify the educational and occupational problems separately. Therefore, we could analyze the prevalence of each of these subtopics. The second reason for creating the lexicon was to verify the result from the LDA. Lexicons about coronavirus and pandemic development have been created to verify the results from the LDA. For example, according to the randomly selected sample posts in each topic group, posts that include the tokens “no mask” and “without mask” mentioned the worries of being infected by people who do not wear masks on the street. Since the topic annotations with the lexicon approach are more explainable, it was used to verify the LDA. After annotating the topics, the trend on the prevalence of topics with the LDA approach and lexicon approach were visualized for comparison.

According to Figure 7, the results using the 2 approaches were comparable. The Person correlation coefficient of the results for the topic “Fear of coronavirus” from the 2 approaches was 0.995, and the correlation coefficient for the “Pandemic development” topic was 0.829.

The keywords for each of the topics were selected by evaluating the sample posts from the topic groups in the LDA model. As we assumed we had no prior knowledge of which mental health issues were expressed on the platform and the commonly used keywords for each of the topics, the LDA model was applied before using the lexical approach. Once the keywords and topics were obtained, we could use the lexical approach to label the topics mentioned in each of the posts and visualize the trends. In some cases, the posts may describe the topics without using the keywords. For example, in the topic “mental health symptoms” in the LDA model, the text in a post may include the words “feel,” “anxious,” “depressed,” and “tired”; however, those words are also common in other topics. Therefore, it may be unsuitable to use those keywords to identify the topic “mental health symptoms” using the lexical approach. For this case, the LDA model is needed to identify that topic. In this study, both methods were used, showing similar results for the topics “Fear of coronavirus” and “Pandemic development” (measured using the correlation coefficient).

We also plotted the trends in the number of posts for each topic separately in Table 2. The keywords for each of the topic groups were obtained by evaluating the sample posts. If a post contained words included in the lexicon, the post was assumed to include the content of the corresponding topic. Each post could include content on more than 1 topic. To observe relationships with the development of the pandemic, the total number of COVID-19 cases and the vaccinated population size were included in the same figure.

Figure 8 shows the result of the lexical approach for each of the topics. According to Figure 8, the number of posts mentioning “education problems” had a declining trend from March 2020 to June 2020. The peak, in March 2020, could be due to the

start of school closures, when both students and teachers were not familiar with online learning. After that, the students may have adapted to the situation. The number of posts with content regarding online learning may have declined when the summer holidays approached. However, the number of posts with the topic of "educational problems" increased from June 2020 to October 2020. This may suggest that the prevalence of education-related stress increased with time. Then, the number of related posts decreased, starting from October 2020. Compared with other topics, the prevalence of education-related discussions started decreasing before the vaccine distribution. The trend was independent of the number of cases.

To understand which stressor was the most prevalent at different stages of the pandemic, we measured the percentage of posts containing the topics in each month and plotted the trend.

According to Figure 9, in September 2020, more than 40% of the posts mentioned loneliness. After September 2020, there was a decreasing trend for the topic of loneliness.

Table 2. Lexicon for COVID-19 stressors.

Topics	Tokens
Education problems	college, online learning, class, semester, freshman
Occupational problems	lost job, unemployed, laid off, income, money, quit job, career
Lonely	social interaction, interact, connection, lonely, friendless, feel alone, loneliness, friendless, social life, friendship, socialize, make friends, new friends, disconnected
Fear of coronavirus	no mask, without mask, maskless, unmasked, grocery, panic, precautions, coworker, cough, exposed, wash, temperature, OCD
Pandemic development	forever, permanent, back normal, new normal, ever end, never ending, endless, lose hope, normal life

Figure 7. This figure compares the trends in the prevalence of topics between the latent Dirichlet allocation (LDA) model (solid lines) and the lexicon (dashed lines).

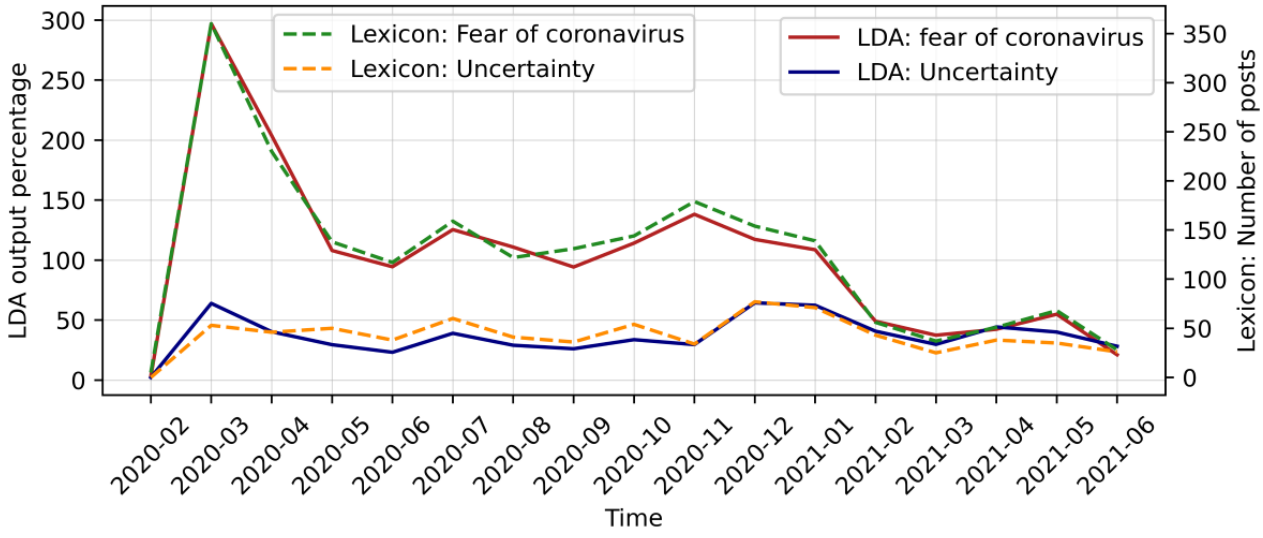


Figure 8. Trend on the number of posts mentioning each of the topics in the lexical approach: (A) educational problems, (B) occupational problems, (C) lonely, (D) fear of coronavirus, (E) pandemic development.

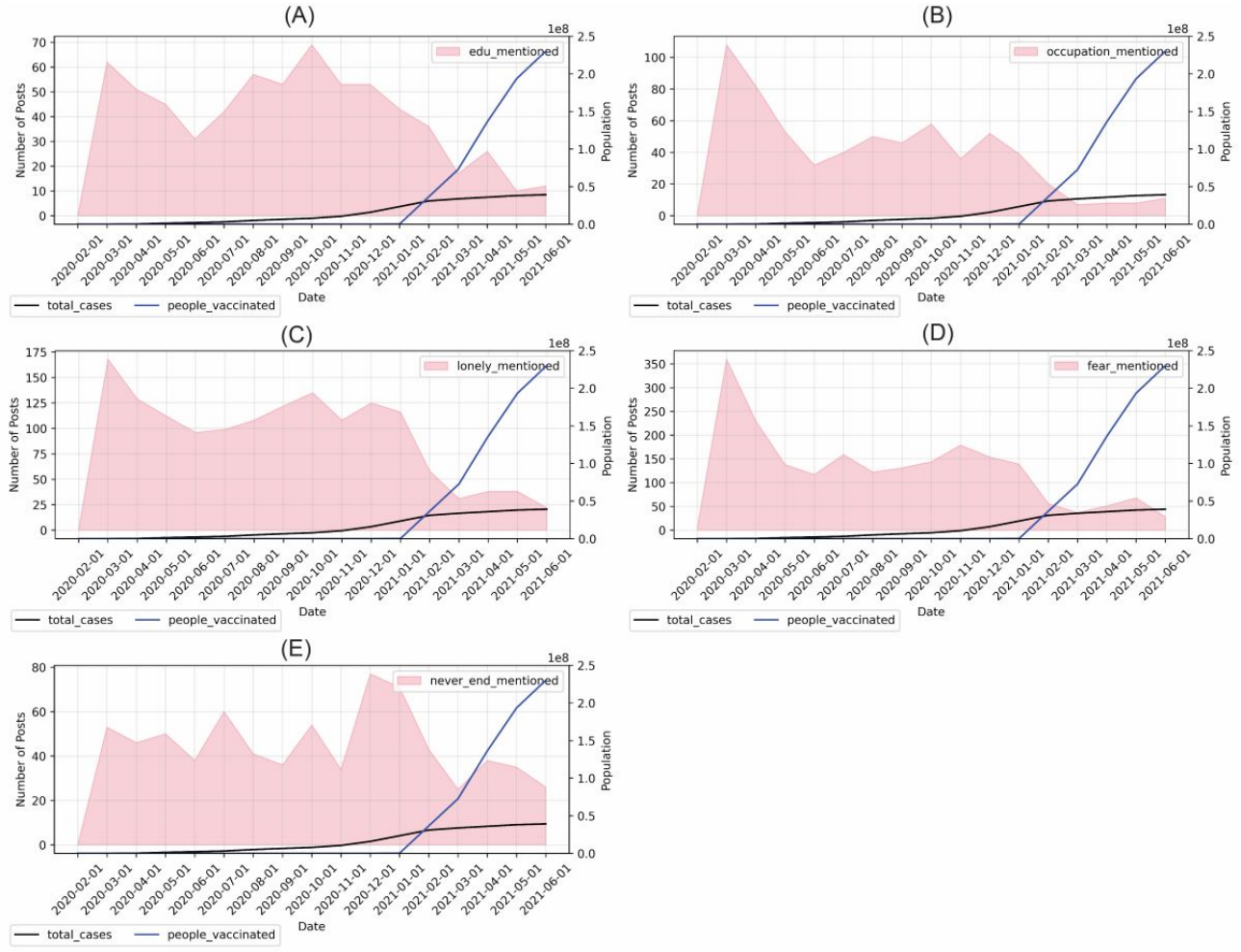
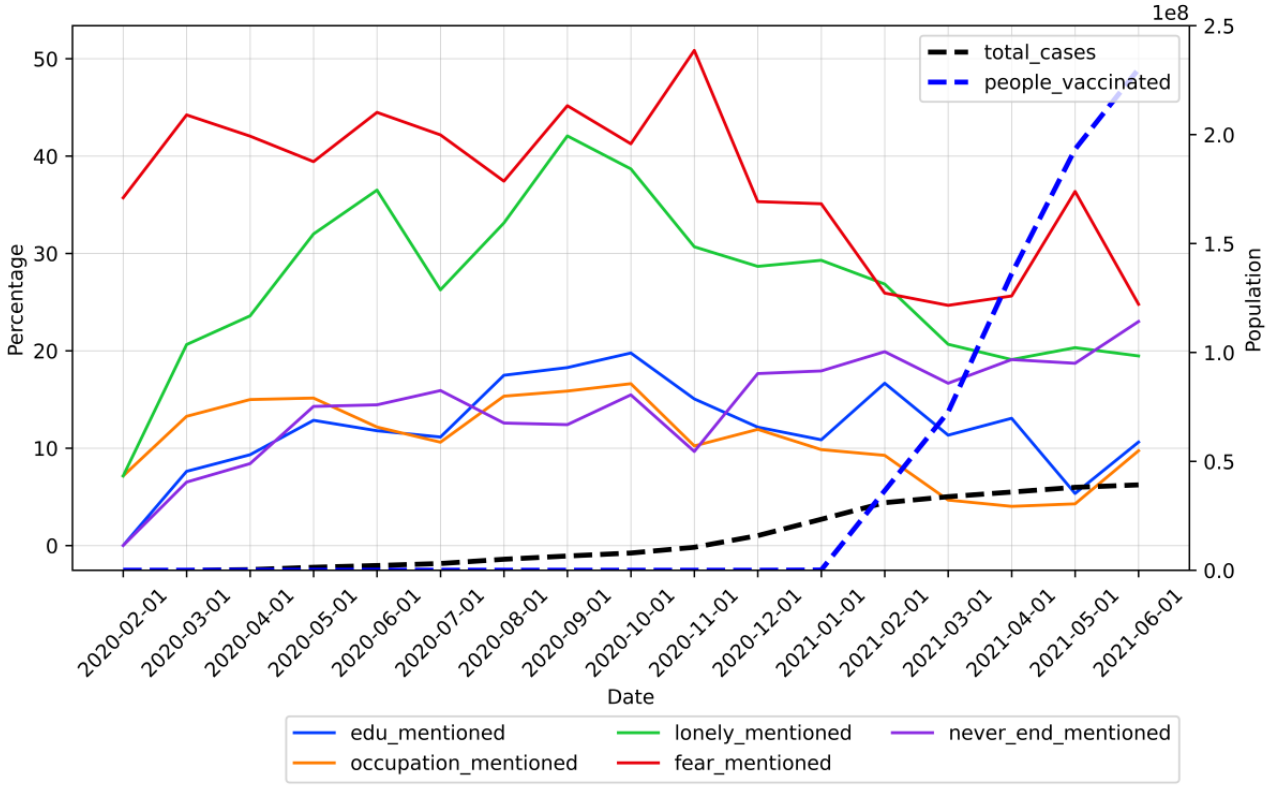


Figure 9. Percentage of posts mentioning each of the topics in each month (dominant topic).



Discussion

Principal Findings

In this study, with the use of the topic model on Reddit data (subreddit \rCOVID19_support), we identified 6 topics related to the pandemic, which were "fear of coronavirus," "educational and occupational problems," "family problems," "problems related to social relationships," "mental health symptoms," and "uncertainty about development of pandemic."

According to the result of our study, the prevalence of discussions on "fear of coronavirus" dropped significantly after the start of vaccination. One of the possible explanations is that the increase in the vaccinated population may have reduced the perceived risk of COVID-19. Perez-Arce et al [20] suggested the uptake of vaccines has potential benefits on mental health because it reduces the perceived risks of infection; people who are recently vaccinated are less worried about being infected and would engage in social activities that they considered as conveying a high risk of infection before. In the study, it was also suggested that the unvaccinated population could be less worried about being infected or infecting family members as more people became vaccinated. This could explain the significant decrease in the prevalence of discussions on "fear of coronavirus" and also the topics related to family problems. In the LDA topic group about family, there were posts related to the worries of infecting family members, struggles about whether one should avoid visiting grandparents, and arguments with family members due to different opinions about the balance between social distancing and social activities. These stressors may have been alleviated by the decrease in perceived risks of infection when the vaccinated population was increasing.

The number of posts about "uncertainty about pandemic development" did not have a notable drop while the vaccinated population was growing, but the trend shows a correlation with the number of cases. People may have been uncertain about the length of lockdown and how long social distancing requirements would last. This could be correlated with the number of cases instead of the perceived risks of infections. Briscese et al [21] suggested that, when the length of social distancing measures is extended, the population may think the goal is unachievable and feel frustrated. In our study, the prevalence of discussions about uncertainty significantly rose in April 2021, when the Delta variant was becoming prevalent in the United States. This trend may reflect the frustration of knowing the length of social distancing measures would need to be extended due to the new variant.

Comparison With Prior Research

In our study, the LDA models identified some of the COVID-19–related stressors proposed in prior studies. Taylor et al [22] devised the "COVID Stress Scale" to measure COVID-19–related distress and identify people in need of mental health services. The scale includes symptoms such as "danger and contamination fears" and "compulsive checking and reassurance seeking." The study found that some people may have worries such as "social distancing is not enough to keep me safe from the virus," "people around me will infect me with the virus," and "can't keep my family safe from the virus." In our data set, those worries were expressed in the LDA topic group "fear of coronavirus." In the research by Park et al [11], the most common COVID-19–related stressors in April 2020 were "reading/hearing about the severity and contagiousness of COVID-19" and "uncertainty about length of quarantine and

social distancing requirements." Those stressors were also identified by the LDA model applied to the \rCOVID19_support posts in our study. This verifies the potential of using this subreddit to capture the trend of major COVID-19–related stressors.

The trend observed in our study is consistent with those of prior studies. Yarrington et al [1] studied the changes in sentiment such as anxiety, tiredness, and depression during the different stages of the COVID-19 pandemic, with the use of data collected by a mental health app in the United States. The results of the study showed that the anxiety level was the highest in the pre stage (February 2, 2020 to March 11, 2020) and then decreased and remained stable during the acute stage (March 12 2020 to April 15, 2020) and sustained stage (April 16, 2020 to July 6, 2020) of stay-at-home orders. Daly and Robinson [23] assessed the psychological distress in the United States using the Patient Health Questionnaire-4 (PHQ-4). Distress levels also declined from April 2020 to June 2020. In our study, the number of posts mentioning the fear of coronavirus was the highest in March 2020, dropped in April 2020 and May 2020, and then remained stable from May 2020 to November 2020.

Limitations

In this study, the prevalence of stressors was only compared with the number of cases and the vaccinated population and not compared with specific safety measures in specific cities. This was due to the anonymity on Reddit. The demographic information of users was unknown, and the data set obtained in this study included posts written by users in different countries. Every city has implemented social distancing measures and lockdowns at different times, depending on the number of cases and the hospitalization rates. Due to this situation, we could not analyze the relationship between mental health status and the safety measures.

Similar to other studies that utilized social media data, the data extracted could only represent the population who would share their experience and feelings on social media. During the pandemic, children and older adults were strongly impacted by the changes, but it is unlikely that they shared their experiences on Reddit or other social media. To understand the needs of people with different demographics, questionnaires or interview-based studies are still required.

Implications for Public Health and Future Studies

According to the results of this study, the stressors that were caused by perceived risks were alleviated since the beginning of the vaccine distribution. However, the stressors related to the frustration of uncertainty on the length of social distancing measures then became the major stressors. Lockdown and social distancing policies may depend on the hospitalization rate, transmissivity, and severity of the virus. With a high proportion of the population vaccinated and more experience with handling COVID-19 patients, lockdowns such as those that occurred in the first 2 waves of COVID-19 are not likely to be required. In terms of alleviating mental distress, the government may consider explaining to the public that the health care system is prepared to handle new outbreaks of coronavirus and we are in the process of getting back to normal life.

With the use of the lexicons created in this study, we obtained the posts or the sentences that described the worries of getting infected and the uncertainty about whether the pandemic is never-ending. This could be used as a data set for training machine learning classifier models to detect tweets that describe the stressors. Unlike Reddit posts, geographic information is available for tweets. By specifying the users' location of tweets, we could analyze the mental distress impacts of specific social distancing policies. We could also establish time series models to quantify and predict the expected effects on stressors. This could help policy makers to estimate the impacts on mental distress before implementing a policy.

In 2022, a new variant, Omicron, became prevalent, and there were updates to social distancing measures. In the future, we can use similar methods in this study to compare the trends in stressors at the time of the Delta variant and Omicron variant.

Conclusions

In this study, we applied topic modeling to a data set that contained Reddit posts in \rCOVID19_support to identify the COVID-19–related psychosocial stressors and to visualize the trend in the prevalence of the stressors. Compared with existing research, which utilized NLP techniques on social media data to study the mental health impacts of the pandemic, our study focused on stressors instead of mental health status. The data set used in this study included posts that were created in a time period of more than 1 year during the pandemic. This allowed us to compare the difference in the prevalence of stressors before and after vaccines were distributed. The proposed topic model also allowed for monitoring the dominant stressors, which enables mental health support providers to notice changes in stressors at different stages of the pandemic. This study demonstrated the potential of using topic modeling on social media discussions to identify event-specific stressors and create a dashboard to analyze and monitor the trends. We hope the findings in this study will provide insights for health care providers and social workers to address the needs of COVID-19–related mental health support. Furthermore, we hope the NLP techniques used in this study will be applied to analyze psychosocial stressors and create corresponding lexicons of future events such as pandemics, protests, or financial crises.

Conflicts of Interest

None declared.

References

1. Yarrington JS, Lasser J, Garcia D, Vargas JH, Couto DD, Marafon T, et al. Impact of the COVID-19 pandemic on mental health among 157,213 Americans. Journal of Affective Disorders 2021 May 01;286:64-70. [doi: 10.1016/j.jad.2021.02.056] [Medline: 33677184]
2. De Choudhury M, Counts S, Horvitz E. Social Media as a Measurement Tool of Depression in Populations. 2013 Presented at: 5th Annual ACM Web Science Conference; May 2-4, 2013; Paris, France p. 47-56. [doi: 10.1145/2464464.2464480]
3. Larsen ME, Boonstra TW, Batterham PJ, ODea B, Paris C, Christensen H. We feel: mapping emotion on Twitter. IEEE J Biomed Health Inform 2015 Jul;19(4):1246-1252. [doi: 10.1109/JBHI.2015.2403839] [Medline: 25700477]
4. Medford RJ, Saleh SN, Sumarsono A, Perl TM, Lehmann CU. An "Infodemic": leveraging high-volume Twitter data to understand early public sentiment for the coronavirus disease 2019 outbreak. Open Forum Infectious Diseases 2020 Jul;7(7):ofaa258. [doi: 10.1093/ofid/ofaa258] [Medline: 33117854]
5. Jelodar H, Wang Y, Orji R, Huang S. Deep sentiment classification and topic discovery on novel coronavirus or COVID-19 online discussions: NLP using LSTM recurrent neural network approach. IEEE J Biomed Health Inform 2020 Oct;24(10):2733-2742. [doi: 10.1109/JBHI.2020.3001216] [Medline: 32750931]
6. Jang H, Rempel E, Roth D, Carenini G, Janjua NZ. Tracking COVID-19 discourse on Twitter in North America: infodemiology study using topic modeling and aspect-based sentiment analysis. J Med Internet Res 2021 Feb 10;23(2):e25431 [FREE Full text] [doi: 10.2196/25431] [Medline: 33497352]
7. Biester L, Matton K, Rajendran J, Provost EM, Mihalcea R. Quantifying the Effects of COVID-19 on Mental Health Support Forums. 2020 Presented at: Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020; 16th – 20th November 2020; online URL: http://arxiv.org/abs/2009.04008 [doi: 10.18653/v1/2020.nlpcovid19-2.8]
8. Low DM, Rumker L, Talkar T, Torous J, Cecchi G, Ghosh SS. Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on Reddit during COVID-19: observational study. J Med Internet Res 2020 Oct 12;22(10):e22635 [FREE Full text] [doi: 10.2196/22635] [Medline: 32936777]
9. APA dictionary of psychology: psychosocial stressor. American Psychological Association. URL: https://dictionary.apa.org/psychosocial-stressor [accessed 2021-11-24]
10. Mowery D, Bryan C, Conway M. Towards Developing an Annotation Scheme for Depressive Disorder Symptoms: A Preliminary Study using Twitter Data. 2015 Presented at: 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; 2015; Denver, CO. [doi: 10.3115/v1/w15-1211]
11. Park CL, Russell BS, Fendrich M, Finkelstein-Fox L, Hutchison M, Becker J. Americans' COVID-19 stress, coping, and adherence to CDC guidelines. Journal of General Internal Medicine 2020 Aug 29;35(8):2296-2303 [FREE Full text] [doi: 10.1007/s11606-020-05898-9] [Medline: 32472486]
12. De Choudhury M, De S. Mental health discourse on reddit: self-disclosure, social support, and anonymity. Proceedings of the International AAAI Conference on Web and Social Media 2014;8(1):71-80 [FREE Full text]
13. Naseem SS, Kumar D, Parsa MS, Golab L. Text Mining of COVID-19 Discussions on Reddit. 2021 Presented at: International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT); December 14-17, 2020; Melbourne, Australia. [doi: 10.1109/wiiat50758.2020.00104]
14. pushshift/api. GITHUB. URL: https://github.com/pushshift/api [accessed 2021-10-24]
15. Vajjala S, Majumder B, Gupta A, Surana H. Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems. Newton, MA: O'Reilly Media; 2020.
16. Boyd-Graber L, Hu Y, Mimno D. Applications of topic models, vol. 11. Boston, MA: Now Publishers Incorporated; 2017.
17. Ritchie H, Mathieu E, Rodés-Guirao L, Appel C, Giattino C, Ortiz-Ospina E, et al. Coronavirus Pandemic (COVID-19). Our World in Data. 2020. URL: https://ourworldindata.org/coronavirus [accessed 2022-06-24]
18. Clement J. Regional distribution of desktop traffic to Reddit.com as of February 2022 by country. Statista. 2022 May 03. URL: https://www.statista.com/statistics/325144/reddit-global-active-user-distribution/ [accessed 2022-06-24]
19. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. World Health Organization. 2020 Mar 11. URL: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2022-06-24]
20. Perez-Arce F, Angrisani M, Bennett D, Darling J, Kapteyn A, Thomas K. COVID-19 vaccines and mental distress. PLoS One 2021 Sep 8;16(9):e0256406 [FREE Full text] [doi: 10.1371/journal.pone.0256406] [Medline: 34496006]
21. Briscese G, Lacetera N, Macis M, Tonin M. Expectations, reference points, and compliance with COVID-19 social distancing measures. National Bureau of Economic Research. 2020. URL: https://www.nber.org/papers/w26916 [accessed 2022-06-24]
22. Taylor S, Landry CA, Paluszek MM, Fergus TA, McKay D, Asmundson GJ. Development and initial validation of the COVID Stress Scales. J Anxiety Disord 2020 May;72:102232 [FREE Full text] [doi: 10.1016/j.janxdis.2020.102232] [Medline: 32408047]
23. Daly M, Robinson E. Psychological distress and adaptation to the COVID-19 crisis in the United States. J Psychiatr Res 2021 Apr;136:603-609 [FREE Full text] [doi: 10.1016/j.jpsychires.2020.10.035] [Medline: 33138985]

Abbreviations

LDA: latent Dirichlet allocation

NLP: natural language processing
PHQ-4: Patient Health Questionnaire 4
t-SNE: t-distributed Stochastic Neighbor Embedding
TF-IDF: term frequency-inverse document frequency

Edited by M Gisondi, A Raja; submitted 08.02.22; peer-reviewed by M Lotto, N Yu, A Joseph, K Uludag; comments to author 12.04.22; revised version received 29.04.22; accepted 21.06.22; published 13.07.22.

Please cite as:
Leung YT, Khalvati F
Exploring COVID-19–Related Stressors: Topic Modeling Study
J Med Internet Res 2022;24(7):e37142
URL: https://www.jmir.org/2022/7/e37142
doi:10.2196/37142
PMID:35731966

Original Paper

The Relationship Between Information Sources, Health Literacy, and COVID-19 Knowledge in the COVID-19 Infodemic: Cross-sectional Online Study in Japan

Mami Inoue[1], BS; Kanako Shimoura[1], PhD; Momoko Nagai-Tanima[1], PhD; Tomoki Aoyama[1], MD, PhD

Department of Human Health Sciences, Graduate School of Medicine, Kyoto University, Kyoto, Japan

Corresponding Author:
Momoko Nagai-Tanima, PhD
Department of Human Health Sciences
Graduate School of Medicine
Kyoto University
53, Kawahara-cho
Shogoin Sakyo-ku
Kyoto, 606-8507
Japan
Phone: 81 075 751 3964
Email: tanima.momoko.8s@kyoto-u.ac.jp

Abstract

Background: The COVID-19 pandemic has caused not only a disease epidemic but also an infodemic. Due to the increased use of the internet and social media, along with the development of communication technology, information has spread faster and farther during the COVID-19 infodemic. Moreover, the increased choice of information sources has made it more difficult to make sound decisions regarding information. Although social media is the most common source of misinformation, other forms of media can also spread misinformation. However, the media sources used by people with high health literacy and COVID-19 knowledge to obtain information are unclear. Furthermore, the association between the use of multiple information sources and health literacy or COVID-19 knowledge is ill-defined.

Objective: This study aims to examine the following 3 aspects regarding the COVID-19 infodemic: (1) the relationship between health literacy, COVID-19 knowledge, and the number of information sources used; (2) the impact of media use on health literacy; and (3) the impact of media use on COVID-19 knowledge.

Methods: An online cross-sectional study was conducted in November 2021. Participants were 477 individuals aged 20-69 years. After obtaining consent to participate in the study, participants were asked about sociodemographic indicators, sources of health-related information, health literacy, and COVID-19 knowledge. Sources of health-related information were categorized into 4 types: mass media, digital media, social media, and face-to-face communication. The Spearman rank correlation test was conducted to determine the relationship between health literacy, the number of correct answers to COVID-19 knowledge, and the number of information sources used. Multiple regression analysis was conducted with health literacy and the number of correct answers as dependent variables, the 4 media types as independent variables, and age and sex as adjustment variables.

Results: Mass media was the most frequently used source of information, followed by digital media, face-to-face communication, and social media. Social media use was significantly higher among individuals aged 20-29 years than among other age groups. Significant positive correlations were found between health literacy, the number of positive responses to COVID-19 knowledge, and the number of information sources used. Multiple linear regression analysis showed that health literacy is associated with access to information from digital media and face-to-face communication. Additionally, COVID-19 knowledge was associated with access to information from mass media, digital media, and face-to-face communication.

Conclusions: Health literacy and COVID-19 knowledge could be improved using diverse information sources, especially by providing opportunities to use digital media and face-to-face communication. Furthermore, it may be important to improve health literacy and provide accurate knowledge about COVID-19 to young adults.

(J Med Internet Res 2022;24(7):e38332) doi:10.2196/38332

KEYWORDS

COVID-19 infodemic; information source; health literacy; COVID-19 knowledge; social media; cross-sectional study; mass media; digital media

Introduction

The World Health Organization (WHO) declared COVID-19 as an infodemic at the Munich Security Conference in February 2020 [1]. An infodemic is a situation in which people are confused owing to a large amount of misinformation or false information during a disease outbreak [2]. The spread of misinformation during disease outbreaks occurred even during the Middle Ages [3]. However, this phenomenon is now amplified by the development of internet communication technology and the widespread use of social media. During the COVID-19 pandemic, TikTok reported a 38% increase in users, while Facebook and Twitter reported an almost 8% increase in users [4]. Consequently, the information spread much faster and farther away. Fake news and misinformation regarding COVID-19 have varied and have been confirmed multiple times [5]. For example, conspiracy theories state that COVID-19 is being spread by 5G communication technology and that the pandemic is an elaborate hoax spread mainly through Twitter [6,7]. Previous studies have reported that misinformation about COVID-19 spread rapidly on social media, including Facebook and YouTube [8]. Furthermore, a study examining the spread of 126,000 untruthful news stories over an 11-year period via Twitter found that misinformation was shared 70% more often than true information [9]. Social media is considered a prime source of misinformation because anyone, not just experts, can disseminate information.

However, examples of misinformation spread through sources other than social media have also been reported, such as the fake news about the shortage of toilet paper due to the COVID-19 pandemic since most toilet papers are made in China. This news spread in Japan in February and March 2020 [10]. When asked where they received this information, nearly 60% of the participants stated that television (TV) was the most common source [11]. In fact, the source of this fake news was social media, but it had not spread that far. However, it spread rapidly once it was picked up by TV programs and news sites. Therefore, any information source, not just social media, can be considered a potential source of misinformation.

Numerous sources of information are available. Information sources have been identified in previous studies and in the white paper on information and communications in Japan [11-13]. The major sources of information fall into 4 categories: face-to-face communication, such as conversations with family and friends; mass media, such as TV and newspapers; digital media, such as internet searches and news sites; and social media, such as Twitter, Facebook, and YouTube.

Face-to-face communication is a means of accessing information through conversation with others. Previous studies have reported that medical professionals, family members, and friends are sources of information [12-14]. Most people obtain health-related information from health care professionals [12-14]. Mass media is the traditional method of accessing information. Mass media refers to media that conveys information in a public, indirect, or 1-way manner, such as TV, radio, newspapers, magazines, and public relations materials. However, with the development of communication technology, the use of mass media has declined. In a 2014 survey, 70% of participants aged between 10 and 69 years obtained information through mass media [15]. However, in a 2021 survey, access to information through mass media decreased to less than 50% [11]. In the place of mass media, access to information through the internet is gaining ground. Digitized information obtained through the internet is referred to as digital media. Examples include the use of search engines, browsing webpages, and applications. Digital media also include social media. Social media is described as information that can be easily transmitted and exchanged or content that can be created and exchanged by anyone using the internet [16,17]. Social media is a digital medium that allows 2-way communication between individuals. Because people access information through a combination of face-to-face communication, mass media, digital media, and social media, determining the true information is considered more complex.

Health literacy is important in determining true information, especially in the context of infodemics [18,19]. Health literacy is the ability to access, understand, evaluate, and use information and services to promote and maintain health and well-being [20].

A previous study conducted in Australia reported that people with low health literacy have more difficulty in finding and understanding information about COVID-19 than those with high health literacy [21]. Health literacy has been reported to be positively correlated with the frequency of information-seeking behavior and the number of information sources used [22-24]. Thus, people with high health literacy are likely to have a higher frequency of information-seeking behavior, obtain information from multiple information media, and thus have higher disease knowledge. However, the relationship between health literacy, COVID-19 knowledge, and the number and types of information sources used has not been examined.

A previous study examined the use of 25 different information sources and found that highly health-literate people use medical websites and are less likely to use TV, social media, and blogs [12]. However, these 25 sources were too fragmented to examine the relationship between health literacy and multiple information sources. Another study of parents of children with asthma examined the relationship between 5 types of information sources: health professionals, family and friends, the internet, nonprint media, and print media. It was found that individuals with high health literacy obtain information from family, friends, and the internet [23]. However, the number of information sources was small and not exhaustive. Therefore, we thought it might be helpful to categorize information sources into 4 types in order to better capture the relationship and importance of

multiple media to health literacy. The 4 types of media include face-to-face communication, after obtaining an exhaustive list of information sources. Furthermore, as mentioned earlier, misinformation about COVID-19 often originates from social media but may spread through other media. Therefore, it is essential to determine which media helps individuals to understand the information appropriately. Previous studies have examined COVID-19 knowledge and information sources (including 4 types of media) among university students in Jordan and Germany; however, knowledge and information sources were considered separately, and the relationship between them was not identified [22,25]. Several studies have examined the relationship between social media use and COVID-19 knowledge. A study conducted in China examined the relationship between social media use, eHealth literacy, and COVID-19 knowledge and found a positive correlation [26]. However, a study conducted in the United States reported that social media use is positively correlated with trust in misinformation about COVID-19 [27]. Furthermore, a study conducted in Canada reported that social media users misinterpret information about COVID-19 more frequently on social media and traditional news sites [28]; therefore, a unified view has not been reached. Moreover, these studies were limited to information obtained from social media and the internet and did not examine other information sources.

Therefore, this study aims to examine the following during the COVID-19 infodemic:

- The relationship between health literacy, COVID-19 knowledge, and the number of information sources used
- The influence of media use on health literacy
- The influence of media use on COVID-19 knowledge

The results may be used to indicate ways to spread true information to a larger population to manage the infodemic.

Methods

Study Design and Recruitment

A cross-sectional online survey was conducted from November 1 to 5, 2021, among individuals aged 20–69 years. Participants were recruited online by Surveroid (Marketing Applications Inc.). The number of participants by sex and age (20-29, 30-39, 40-49, 50-59, and 60-69 years) was set to be the same, and responses were accepted in the order of receipt. The purpose of the study was explained at the beginning of the online questionnaire survey. The submission of the online questionnaire implied consent to participate in the study. The online questionnaires were collected in randomized identification format without asking for personal information, such as names or email addresses. Participants received a reward upon completion based on their registration status in the Surveroid database. A survey request was sent to 8809 individuals via email. A total of 512 (5.8%) responses were received over 5 days of recruitment. Of these, 35 (6.3%) responded that they did not obtain health-related information and were therefore excluded from the study. The final number of participants was 477 (5.4%).

Ethical Considerations

The Medical Ethics Committee of Kyoto University, Japan, approved this study (#R3215).

Measures

The online questionnaire included 4 components or groups of items, which required 5 minutes for completion: (1) sociodemographic indicators and experiences during the COVID-19 pandemic, (2) sources of health-related information, (3) health literacy, and (4) COVID-19 knowledge questions.

Sociodemographic Indicators

Participants were asked about their sex, age, and education level.

Sources of Health-Related Information

Participants were provided with a list of 13 information sources in multiple-response format. They were asked to select the sources they regularly used to obtain health-related information. The list of 13 information sources was compiled from previous studies conducted in Japan and the items used in surveys conducted by the Ministry of Internal Affairs and Communications [11,12,24]. The 13 information sources are listed in Textbox 1.

Textbox 1. List of 13 information sources.

1. Television (TV)
2. Radio
3. Newspaper
4. Publications (eg, magazines)
5. Municipal newsletters
6. Websites (eg, government and medical manufactures)
7. Web search
8. News apps
9. Video sites (eg, YouTube)
10. Social networking services (SNSs, eg, Twitter, Instagram, and Facebook)
11. Hospitals and pharmacies
12. Family
13. Friends

Health Literacy

Health literacy was assessed using the Communicative and Critical Health Literacy (CCHL) scale developed by Ishikawa et al [29]. Health literacy comprises 3 components: functional, interactive, and critical [30]. Functional literacy refers to basic reading and writing skills. Interactive literacy refers to advanced cognitive and literacy skills that can be used to actively participate in daily life, extract information from various forms of communication, understand the meaning, and apply new information to changing situations. Critical literacy refers to more advanced cognitive abilities that can be applied to critically analyze and apply information to successfully control a situation. The CCHL is a self-administered questionnaire that evaluates communicative and critical health literacy among the public. It consists of 5 questions rated on a 5-point scale from 1 (strongly disagree) to 5 (strongly agree). The mean score for all questions is calculated, with a higher mean score indicating higher ability. In a previous study conducted in Japan using the CCHL, the mean score was reported to be 3.58-3.7 [29,31].

COVID-19 Knowledge Questions

Participants were asked whether the information about COVID-19 presented in the questions was correct or incorrect. The questions were based on a selection of misinformation that was prevalent in Japan and clearly listed as incorrect on the question and answer (Q&A) page related to COVID-19 provided by the Ministry of Health, Labor and Welfare [32,33]. Participants responded to the questions as correct, unknown, or incorrect. The total number of correct answers was counted as the correct answer score (CAS). Higher scores indicated a higher number of correct answers and greater knowledge of COVID-19, as presented in this study (Table 1).

Table 1. COVID-19 knowledge questions.

Question	Answer
COVID-19 is vulnerable to heat, and low-temperature water (25-35°C) has a bactericidal effect.	Incorrect
Alcohol disinfection is effective against COVID-19.	Correct
COVID-19 vaccine makes you infertile.	Incorrect
Vaccination can lead to infection with COVID-19.	Incorrect
The vaccine can be given during pregnancy, during lactation, or while planning a pregnancy.	Correct
If a vaccinated person becomes infected with a mutated virus, they are likely to become seriously ill.	Incorrect

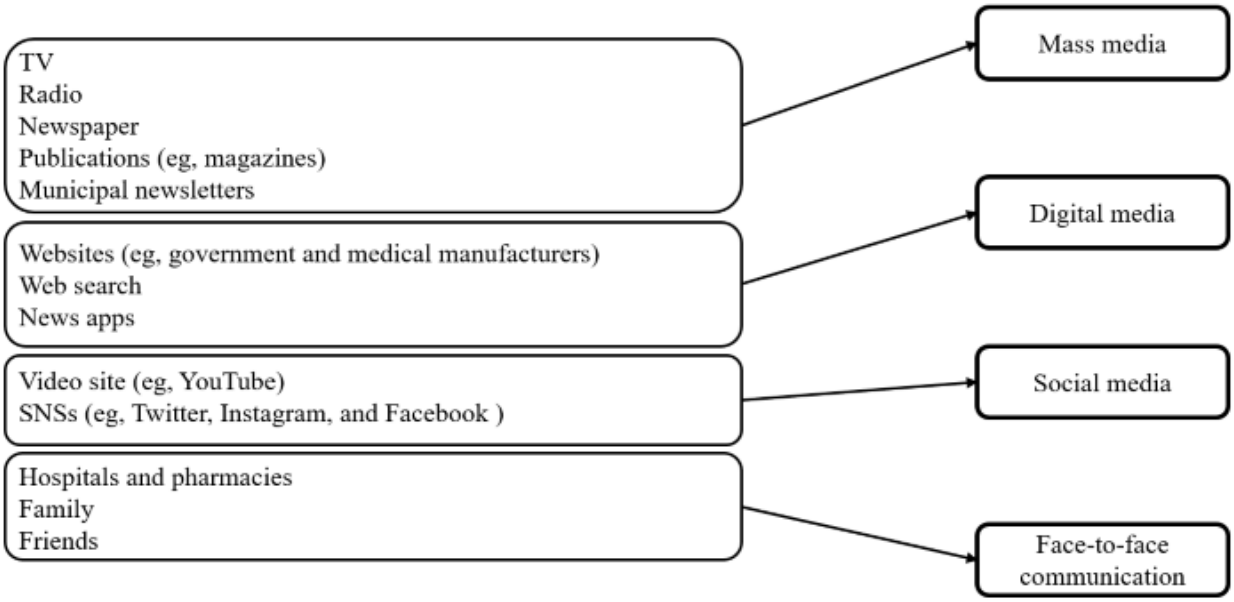
Statistical Analysis

Participant characteristics were analyzed using descriptive statistics. Participant characteristics and the CCHL score or the CAS were compared using Wilcoxon and Kruskal-Wallis rank sum tests. Concerning information sources, the percentage of participants for each item was calculated (number of responses/total number of participants). The Spearman rank correlation test was used to calculate the correlation coefficient between health literacy, the CAS, and the number of information sources used.

The 13 information sources were categorized into 4 media types: mass, digital, social, and face-to-face communication (Figure 1). For example, if participants used TV, radio, and video sites, they were considered to be using mass media and social media. Chi-square tests were used to compare the percentages of those who selected each medium according to sex and age. Multiple linear regression analysis was performed to examine whether

the any of the 4 media types had an impact on the CCHL score and the CAS. The dependent variables were the CCHL score and the CAS, the independent variables were the four types of media, and the adjusted variables were sex and age. Each medium was set as 1 for use and –1 for no use, and sex was set as 1 for males and –1 for females. When there was a correlation between the dependent variables, multicollinearity (variance inflation factor [VIF]) was examined. There was no multicollinearity if the VIF<10. The significance level for rejection of the null hypothesis was 5%. Statistical analysis was performed using JMP Pro version 15.0 statistical software (SAS Institute Japan Co.).

Figure 1. Method of classifying sources of information. SNS: social networking service; TV: television.



Results

Participants' Characteristics

The mean age of the participants was 44.8 (SD 14.3) years. In the comparison of the CCHL score and the CAS by age group, sex, and education level, there was a significant difference only in the CAS by age group; the CAS in the age group of 60-69 years was significantly higher than that in other age groups (Table 2).

Table 2. Participants' characteristics and comparison of the CCHL[a] score and the CAS[b].

Variables	Participants, n (%)	CCHL score, mean (SD)	CAS, mean (SD)
Total	477 (100)	3.61 (0.67)	3.75 (2.01)
Age (years); CCHL *P*=.51, CAS *P*<.001[c]			
20-29	90 (18.9)	3.57 (0.07)	3.31 (0.20)
30-39	92 (19.3)	3.54 (0.07)	3.55 (0.20)
40-49	98 (20.5)	3.59 (0.07)	3.26 (0.20)
50-59	97 (20.3)	3.67 (0.07)	3.88 (0.20)
60-69	100 (21.0)	3.67 (0.07)	4.67 (0.20)[d]
Sex; CCHL *P*=.46, CAS *P*=.84			
Male	240 (50.0)	3.63 (0.04)	3.74 (0.13)
Female	237 (50.0)	3.59 (0.04)	3.76 (0.13)
Education level; CCHL *P*>=.09, CAS *P*=.30			
Middle school	8 (1.7)	3.28 (0.24)	3.00 (0.71)
High school	150 (31.4)	3.54 (0.05)	3.51 (0.16)
Technical school and junior college	117 (24.5)	3.62 (0.06)	4.00 (0.19)
University	181 (38.0)	3.65 (0.05)	3.77 (0.15)
Graduate school	21 (4.4)	3.91 (0.14)	4.14 (0.44)

[a]CCHL: Communicative and Critical Health Literacy.

[b]CAS: correct answer score.

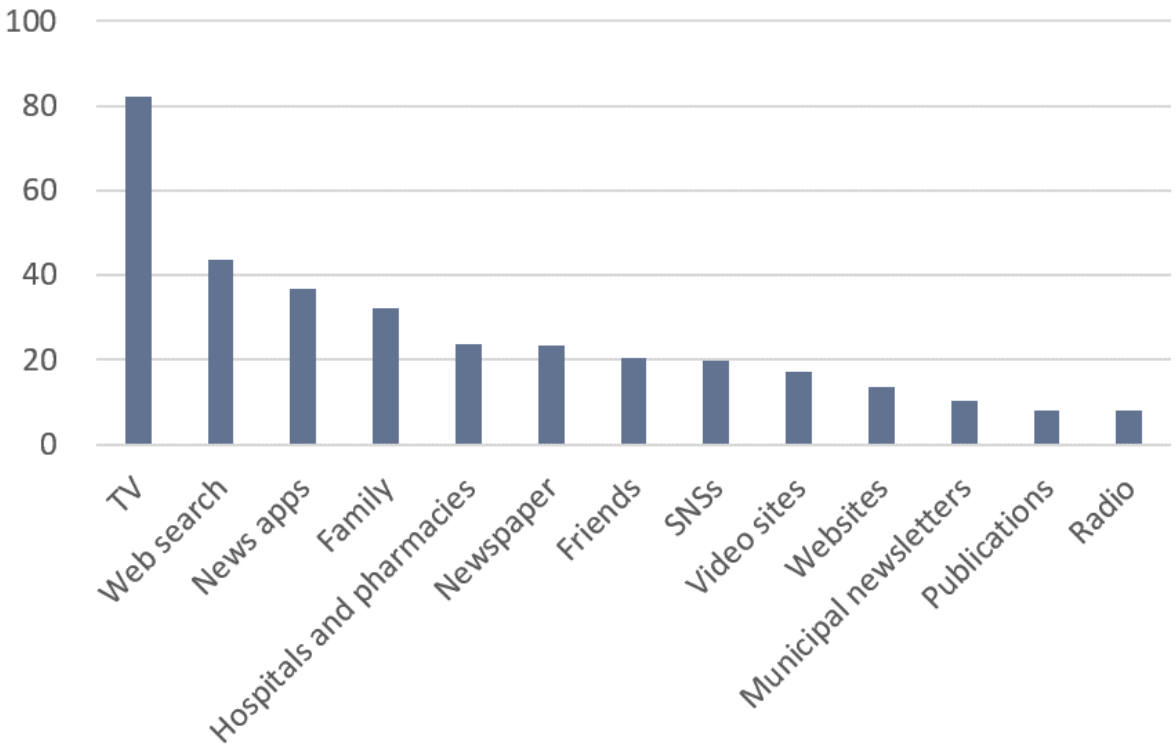
[c]*P*<.05.

[d]The CAS in the age group of 60-69 years was significantly higher than in other age groups.

Percentage of Participants by Information Source

As shown in Figure 2, 392 (82.2%) of the 477 participants received information from the TV. Web searches (n=208, 43.6%) and news apps (n=175, 36.7%) were the next most popular sources. Social networking services (SNSs) and video sites were used by 94 (19.7%) and 82 (17.2%) participants, respectively. Municipal newsletters were used by 49 (10.3%) participants, while publications and radio were used by 39 (8.2%) and 38 (8.0%) participants, respectively.

Figure 2. Percentage of responses for information sources. SNS: social networking service; TV: television.



Correlation Between the CCHL Score, the CAS, and the Number of Information Sources Used

The CCHL score and the CAS were significantly and positively correlated (ρ=0.12, *P*<.001). Additionally, the number of information sources used was significantly and positively correlated with the CCHL score (ρ=0.22, *P*<.001) and the CAS (ρ=0.19, *P*<.001).

Types of Media Associated With Health Literacy and COVID-19 Knowledge

Table 3 shows the percentages for all participants, sexes, and age groups when the selected information sources were categorized into different media types. The largest percentage of participants indicated mass media as their source of information, followed by digital media, face-to-face communication, and social media. There were no differences in the use of mass media, digital media, or face-to-face communication by age group. Social media use was significantly higher among individuals aged 20-29 years than among other age groups (*P*<.001). Table 4 displays the multiple linear regression analysis results, with the CCHL score as the dependent variable, the 4 types of media as independent variables, and sex and age as adjustment variables. Table 5 displays the multiple linear regression analysis results, with the CAS as the dependent variable, the 4 types of media as independent variables, and sex and age as adjustment variables. In all cases, a VIF<10 was observed. The CCHL score was significantly associated with access to information from digital media (β=.14, *P*=.003), where β is the standardized regression coefficient, and face-to-face communication (β=.11, *P*=.02). The CAS was significantly associated with access to information from mass media (β=.09, *P*=.05), digital media (β=.17, *P*<.001), and face-to-face communication (β=.10, *P*=.02).

Table 3. Distribution by media type (N=477).

Media type	Total, n (%)	Males, n (%)	Females, n (%)	*P* value	20-29 years, n (%)	30-39 years, n (%)	40-49 years, n (%)	50-59 years, n (%)	60-69 years, n (%)	*P* value
Mass media	413 (86.6)	206 (85.8)	207 (87.3)	.69	73 (81.1)	78 (84.8)	83 (84.7)	82 (84.5)	97 (97.0)	.42
Digital media	288 (60.4)	144 (60.0)	144 (60.8)	.93	53 (58.9)	55 (59.8)	54 (55.1)	64 (66.0)	62 (62.0)	.78
Social media	144 (30.2)	69 (28.8)	75 (31.7)	.55	50 (55.6)	32 (34.8)	26 (26.5)	16 (16.5)	20 (20.0)	<.001[a,b]
Face-to-face communication	222 (46.5)	95 (39.6)	127 (53.6)	.002[a]	40 (44.4)	39 (42.4)	47 (48.0)	46 (47.4)	50 (50.0)	.73

[a]*P*<.05.

[b]Social media use in the age group of 20-29 years was significantly higher than in the other age groups. Social media use in the age group of 30-39 years was significantly higher than in individuals aged 50-59 and 60-69 years. Individuals aged 40-49 years had significantly higher social media use than individuals aged 50-59 years.

Table 4. Multiple linear regression analysis with the CCHL[a] score as the dependent variable and sex[b] and age as adjustment variables.

Media type	β[c]	95% CI	*P* value
Mass media	.04	–0.04 to 0.13	.33
Digital media	.14	0.03-0.16	.003[d]
Social media	–.02	–0.09 to 0.05	.63
Face-to-face communication	.11	0.01-0.14	.02[d]

[a]CCHL: Communicative and Critical Health Literacy.

[b]Media use and male sex were set as 1.

[c]Standardized regression coefficient.

[d]*P*<.05.

Table 5. Multiple linear regression analysis with the CAS[a] as the dependent variable and sex[b] and age as adjustment variables.

Media type	β[c]	95% CI	*P* value
Mass media	.09	0-0.51	.05[d]
Digital media	.17	0.17-0.55	<.001[d]
Social media	–.02	–0.25 to 0.16	.69
Face-to-face communication	.10	0.03-0.39	.02[d]

[a]CAS: correct answer score.

[b]Media use and male sex were set as 1.

[c]Standardized regression coefficient.

[d]*P*<.05.

Discussion

Principal Findings

This study examined the following 3 aspects regarding the COVID-19 infodemic: (1) the relationship between health literacy and COVID-19 knowledge and the number of information sources used, (2) the influence of media use on health literacy, and (3) the influence of media use on knowledge of COVID-19. To the best of our knowledge, this is the first study to examine whether access to information from the 4 major types of media sources is associated with health literacy and COVID-19 knowledge during the COVID-19 infodemic. The Spearman rank correlation test revealed a significant positive correlation between health literacy, COVID-19 knowledge, and the number of information sources used. Multiple linear regression analysis revealed that those with higher health literacy access information through digital media and face-to-face communication. Additionally, the more COVID-19 knowledge people had, the more they accessed information from mass media, digital media, and face-to-face communication.

The CCHL scores of the participants in this study were comparable to those reported in previous studies conducted in Japan [29,31]. Furthermore, health literacy was positively

correlated with age [34]. In this study, there was no significant difference between age and the CCHL score. However, the higher the age, the higher the CCHL score, indicating a similar trend as in previous studies. The education level tended to be higher than the census results [35]. Thus, the participants in this study may have had a higher level of education than the general Japanese population.

Participants were more likely to access health-related information from mass media. Previous studies conducted outside Japan have reported that the highest percentage of participants used family members and medical professionals, such as primary care providers and nurses, as information sources rather than mass media [12,13,23]. Furthermore, in studies conducted among young adults, the highest percentage of participants used the internet [22,25]. However, studies conducted in Japan have shown that older adults access information from mass media and family members [24] and that younger adults access information from the internet more than older adults [11]. The results of this study showed that TV, web searches, news apps, and family members are the most common information sources, in that order, with medical professionals ranking fifth overall. However, people in Japan may access information from medical professionals less frequently than those outside the country. The average age at which participants accessed information from mass media, digital media, and face-to-face communication was almost the same; however, the average age was 7 years younger for social media. In recent years, the age group using the internet has expanded, with a 2020 survey showing that almost 100% of individuals aged 20-59 years and 80% of individuals aged above 60 years use the internet [11]. However, social media use is more frequent among younger age groups, with over 90% of individuals aged 20-29 years and only about 60% of individuals aged above 60 years using social media [11]. Therefore, the average age of social media users in this study was also considered younger.

Additionally, the results indicated that the higher the health literacy, the greater the COVID-19 knowledge. Prior research has shown that individuals with low health literacy have more difficulty finding and understanding information about COVID-19 than those with high health literacy [21]. This study supports these findings. Moreover, the greater the number of information sources used, the higher the health literacy and COVID-19 knowledge. Prior research has shown that individuals with higher health literacy are more likely to use various information sources [23,24]. Perhaps, individuals may have also gained knowledge about COVID-19 by obtaining information from diverse sources. It is also possible that the more information sources they used, the higher their CCHL scores; the CCHL scale includes an item about whether they obtained information from a variety of information sources.

The study results indicate that the higher the health literacy, the more the information accessed through digital media and face-to-face communication. Previous studies have reported that people with higher health literacy are more likely to access information from websites (especially medical-related websites), family, and friends [12,23]. Therefore, the results of this study are consistent with the previous findings. Digital media includes internet searches and news apps. Information accessed through digital media ranges from highly reliable sources, such as public institutions and medical manufacturers, to a considerable volume of unverified and unreliable sources, such as personal blogs. Previous studies have reported that many web pages appear when individuals search for health information; however, there are gaps in availability, with insufficient or contradictory content [36]. Therefore, to access information through digital media, it is necessary to select the required information from the vast amount of available information using appropriate search terms and to understand and analyze the content. This differs from mass media in that it widely conveys information in a 1-way manner. This process utilizes communicative and critical literacy. Previous research has reported that people with low health literacy underestimate high-quality information and overestimate low-quality information on the web, making it difficult to accurately judge the information [37]. It is more difficult for these individuals to select the necessary information from digital media. Therefore, higher health literacy is associated with digital media use.

Moreover, accessing information through face-to-face communication is consistent with the process of communicative literacy in that information is obtained through communication with various people. Therefore, higher health literacy is associated with accessing information through face-to-face communication.

Additionally, the results indicate that the more COVID-19 knowledge participants had, the more they accessed information from mass media, digital media, and face-to-face communication. Previous studies have shown that the source of fake news and misinformation regarding COVID-19 is often social media [5]. Social media can spread uncertain or low-quality information, which can lead to misinformation [38,39]. Furthermore, social media uses algorithms that link content based on how data are handled and prioritized [40]. Once misinformation is viewed, similar information may appear repeatedly and is assumed to be correct. In contrast, mass and digital media (eg, news sites) have a check system to avoid conveying misinformation and convey true information at the same time as misinformation [11]. Therefore, it may be easier for recipients to judge true information. Notably, social media complies with WHO and global health authorities and provides links to the websites of public institutions during the COVID-19 pandemic and reminds people to access information with a high level of evidence [41]. However, in a survey conducted during the COVID-19 pandemic in Japan, 30% of participants confirmed the authenticity of information they thought was untrue, while 50% did not confirm the authenticity of information [42]. This finding indicates that alerts may not be sufficient to lead people to access evidence-based information. Furthermore, our results suggest that misinformed people may access the media to obtain misinformation on their own. In such cases, directing people to highly evidence-based information may not be sufficient. Therefore, it might be possible to improve the situation by conveying misinformation as well as true information at the same time as mass media and digital media, such as news sites.

Considering the results, both health literacy and COVID-19 knowledge were associated with access to information from digital media and face-to-face communication. Health literacy and COVID-19 knowledge may be improved by providing opportunities to use digital media and face-to-face communication.

Moreover, the study results showed that the younger the age, the less the COVID-19 knowledge and the greater the use of social media. Previous studies have shown that younger people have lower health literacy [34]. Since social media is a major source of misinformation on COVID-19 [5], individuals may disseminate and spread information without proper understanding. Furthermore, it may be important for young adults to improve their health literacy and to be provided with the correct knowledge about COVID-19.

Limitations

This study was conducted only in Japan; thus, further studies are needed to generalize the results to other countries. The study only included participants registered with Surveroid, with a response rate of 5.8%. Therefore, it is necessary to increase the demographics and the number of participants to strengthen the results of this study. Furthermore, this was a cross-sectional study and causal relationships could not be demonstrated. A longitudinal study would need to be conducted to demonstrate a causal relationship. The COVID-19 knowledge questions used in this study were obtained from the Ministry of Health, Labor and Welfare website. Since other misinformation has circulated in Japan [42], it is necessary to examine other COVID-19 knowledge in the future. Finally, the participants were asked about the information sources they did or did not use. The relationship between health literacy and COVID-19 knowledge can be examined in more detail by asking detailed questions about the frequency of access and priorities.

Future Perspectives

The COVID-19 pandemic is still ongoing. Therefore, it is crucial to have adequate access to information about COVID-19.

The prevailing misinformation about COVID-19 is changing with time and the type of prevalent virus. The prevalence of misinformation is also likely to vary from country to country. Therefore, it is necessary to generalize this finding by expanding the area and period in which the survey is conducted. Several health literacy scales exist, in addition to the survey instruments used in this study. In particular, a more detailed assessment of social and digital media use may be possible by measuring eHealth literacy. Furthermore, there is a need to increase the level of evidence by conducting longitudinal studies to investigate the frequency of and changes in media use with interventions to improve health literacy.

Conclusion

This study examined the association between health literacy, COVID-19 knowledge, and information sources during the COVID-19 pandemic in Japan. The results revealed that the higher the health literacy, the more the knowledge about COVID-19, and the more information sources used, the higher the health literacy and the more accurate the COVID-19 knowledge. Individuals with higher health literacy were found to access information through digital media and face-to-face communication, while individuals with more knowledge about COVID-19 accessed information through mass media, digital media, and face-to-face communication. Health literacy and COVID-19 knowledge may be improved using various information sources, especially by providing opportunities to use digital media and face-to-face communication. Furthermore, it may be essential to improve health literacy and provide accurate knowledge about COVID-19 to young individuals. During the ongoing COVID-19 infodemic, it is crucial to determine truthful information and avoid being swayed by misinformation.

Acknowledgments

This research did not receive any specific funding from any funding agency, public, for-profit, or nonprofit organization. We would like to thank all participants who participated in the survey and the laboratory for their valuable advice and assistance. We would also like to thank Editage for assistance with English language editing.

Conflicts of Interest

None declared.

References

1. World Health Organization. Munich Security Conference. URL: https://www.who.int/director-general/speeches/detail/munich-security-conference [accessed 2022-03-09]
2. World Health Organization. Infodemic. URL: https://www.who.int/health-topics/infodemic#tab=tab_1 [accessed 2022-03-09]
3. The Lancet Infectious Diseases. The COVID-19 infodemic. Lancet Infect Dis 2020 Aug;20(8):875. [doi: 10.1016/s1473-3099(20)30565-x]
4. Statista. Social Media Use during COVID-19 Worldwide - Statistics & Facts. URL: https://www.statista.com/topics/7863/social-media-use-during-coronavirus-covid-19-worldwide/#topicHeader__wrapper [accessed 2022-03-09]
5. Naeem SB, Bhatti R, Khan A. An exploration of how fake news is taking over social media and putting public health at risk. Health Info Libr J 2021 Jun 12;38(2):143-149 [FREE Full text] [doi: 10.1111/hir.12320] [Medline: 32657000]

6. Ahmed W, Vidal-Alaball J, Downing J, López Seguí F. COVID-19 and the 5G conspiracy theory: social network analysis of Twitter data. J Med Internet Res 2020 May 06;22(5):e19458 [FREE Full text] [doi: 10.2196/19458] [Medline: 32352383]
7. Ahmed W, López Seguí F, Vidal-Alaball J, Katz MS. COVID-19 and the "film your hospital" conspiracy theory: social network analysis of Twitter data. J Med Internet Res 2020 Oct 05;22(10):e22374 [FREE Full text] [doi: 10.2196/22374] [Medline: 32936771]
8. Rocha YM, de Moura GA, Desidério GA, de Oliveira CH, Lourenço FD, de Figueiredo Nicolete LD. The impact of fake news on social media and its influence on health during the COVID-19 pandemic: a systematic review. Z Gesundh Wiss 2021 Oct 09:1-10 [FREE Full text] [doi: 10.1007/s10389-021-01658-z] [Medline: 34660175]
9. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science 2018 Mar 09;359(6380):1146-1151. [doi: 10.1126/science.aap9559] [Medline: 29590045]
10. Konishi Y, Saito T, Ishikawa T, Kanai H, Igei N. How did Japan cope with COVID-19? Big data and purchasing behavior. Asian Econ Pap 2021;20(1):146-167. [doi: 10.1162/asep_a_00797]
11. Ministry of Internal Affairs and Communications. Information and Communications in Japan. White Paper 2021. URL: https://www.soumu.go.jp/johotsusintokei/whitepaper/eng/WP2021/2021-index.html [accessed 2022-03-09]
12. Chen X, Hay JL, Waters EA, Kiviniemi MT, Biddle C, Schofield E, et al. Health literacy and use and trust in health information. J Health Commun 2018;23(8):724-734 [FREE Full text] [doi: 10.1080/10810730.2018.1511658] [Medline: 30160641]
13. Chen X, Orom H, Hay JL, Waters EA, Schofield E, Li Y, et al. Differences in rural and urban health information access and use. J Rural Health 2019 Jun;35(3):405-417 [FREE Full text] [doi: 10.1111/jrh.12335] [Medline: 30444935]
14. Kelley MS, Su D, Britigan DH. Disparities in health information access: results of a county-wide survey and implications for health communication. Health Commun 2016 Oct 09;31(5):575-582. [doi: 10.1080/10410236.2014.979976] [Medline: 26452300]
15. Ministry of Internal Affairs and Communications. Information and Communications in Japan. White Paper 2014. URL: https://www.soumu.go.jp/johotsusintokei/whitepaper/eng/WP2014/2014-index.html [accessed 2022-03-09]
16. Ministry of Internal Affairs and Communications. Information and Communications in Japan. White Paper 2015. URL: https://www.soumu.go.jp/johotsusintokei/whitepaper/eng/WP2015/2015-index.html [accessed 2022-03-09]
17. Kaplan AM, Haenlein M. Users of the world, unite! The challenges and opportunities of social media. Bus Horiz 2010 Jan;53(1):59-68. [doi: 10.1016/j.bushor.2009.09.003]
18. Abel T, McQueen D. Critical health literacy and the COVID-19 crisis. Health Promot Int 2020 Dec 01;35(6):1612-1613 [FREE Full text] [doi: 10.1093/heapro/daaa040] [Medline: 32239213]
19. Chong YY, Cheng HY, Chan HYL, Chien WT, Wong SYS. COVID-19 pandemic, infodemic and the role of eHealth literacy. Int J Nurs Stud 2020 Aug;108:103644 [FREE Full text] [doi: 10.1016/j.ijnurstu.2020.103644] [Medline: 32447127]
20. Nutbeam D, Muscat D. Health promotion glossary 2021. Health Promot Int 2021 Dec 23;36(6):1578-1598. [doi: 10.1093/heapro/daaa157] [Medline: 33822939]
21. McCaffery K, Dodd R, Cvejic E, Ayrek J, Batcup C, Isautier JMJ, et al. Health literacy and disparities in COVID-19-related knowledge, attitudes, beliefs and behaviours in Australia. Public Health Res Pract 2020 Dec 09;30(4):30342012 [FREE Full text] [doi: 10.17061/phrp30342012] [Medline: 33294907]
22. Schäfer M, Stark B, Werner AM, Tibubos AN, Reichel JL, Pfirrmann D, et al. Health information seeking among university students before and during the corona crisis—findings from Germany. Front Public Health 2020 Jan 25;8:616603 [FREE Full text] [doi: 10.3389/fpubh.2020.616603] [Medline: 33585388]
23. Fagnano M, Halterman JS, Conn KM, Shone LP. Health literacy and sources of health information for caregivers of urban children with asthma. Clin Pediatr (Phila) 2012 Mar;51(3):267-273 [FREE Full text] [doi: 10.1177/0009922811421413] [Medline: 21911409]
24. Kinjo H, Ishi K, Saito T, Nomura N, Hamada A. A survey on how older adults access medical and health information and what kinds of problems they face in accessing it. Jpn J Gerontol ? 2017;39(1):7-20. [doi: 10.34393/rousha.39.1_7]
25. Olaimat AN, Aolymat I, Shahbaz HM, Holley RA. Knowledge and information sources about COVID-19 among university students in Jordan: a cross-sectional study. Front Public Health 2020 May 29;8:254 [FREE Full text] [doi: 10.3389/fpubh.2020.00254] [Medline: 32574314]
26. 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]
27. Su Y. It doesn't take a village to fall for misinformation: social media use, discussion heterogeneity preference, worry of the virus, faith in scientists, and COVID-19-related misinformation beliefs. Telemat Inform 2021 May;58:101547. [doi: 10.1016/j.tele.2020.101547]
28. Bridgman A, Merkley E, Loewen PJ, Owen T, Ruths D, Teichmann L, et al. The causes and consequences of COVID-19 misperceptions: understanding the role of news and social media. HKS Misinfo Rev 2020 Jun 18:1-18. [doi: 10.37016/mr-2020-028]
29. Ishikawa H, Nomura K, Sato M, Yano E. Developing a measure of communicative and critical health literacy: a pilot study of Japanese office workers. Health Promot Int 2008 Sep;23(3):269-274. [doi: 10.1093/heapro/dan017] [Medline: 18515303]

30. Nutbeam D. Health literacy as a public goal: a challenge for contemporary health education and communication strategies into the 21st century. Health Promot Int 2006;15(3):259-267 [FREE Full text] [doi: 10.1093/heapro/15.3.259]
31. Kimura N, Kobayashi T. [Association of health literacy with hypertension, diabetes, and dyslipidemia: a cross-sectional survey of a regional Japanese community]. Nihon Koshu Eisei Zasshi 2020;67(12):871-880 [FREE Full text] [doi: 10.11236/jph.67.12_871] [Medline: 33390371]
32. Ministry Health, Labor and Welfare. Coronavirus (COVID-19). URL: https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000164708_00079.html [accessed 2022-03-09]
33. Ministry Health, Labor and Welfare. COVID-19 Vaccination Information. URL: https://v-sys.mhlw.go.jp/en/ [accessed 2022-03-09]
34. Nakayama K, Osaka W, Togari T, Ishikawa H, Yonekura Y, Sekido A, et al. Comprehensive health literacy in Japan is lower than in Europe: a validated Japanese-language assessment of health literacy. BMC Public Health 2015 May 23;15(1):505 [FREE Full text] [doi: 10.1186/s12889-015-1835-x] [Medline: 26001385]
35. Ministry of Internal Affairs and Communications. Census 2015 (in Japanese). URL: http://www.stat.go.jp/data/kokusei/2015/life.html [accessed 2022-03-09]
36. Berland GK, Elliott MN, Morales LS, Algazy JI, Kravitz RL, Broder MS, et al. Health information on the internet: accessibility, quality, and readability in English and Spanish. JAMA 2001 May 23;285(20):2612-2621 [FREE Full text] [doi: 10.1001/jama.285.20.2612] [Medline: 11368735]
37. Diviani N, van den Putte B, Giani S, van Weert JC. Low health literacy and evaluation of online health information: a systematic review of the literature. J Med Internet Res 2015 May 07;17(5):e112 [FREE Full text] [doi: 10.2196/jmir.4018] [Medline: 25953147]
38. Patrick M, Venkatesh RD, Stukus DR. Social media and its impact on health care. Ann Allergy Asthma Immunol 2022 Feb;128(2):139-145. [doi: 10.1016/j.anai.2021.09.014] [Medline: 34555532]
39. 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]
40. Hamamreh R, Awad S. Tag ranking multi-agent semantic social networks. 2017 Presented at: International Conference on Computational Science and Computational Intelligence (CSCI); December 5-7, 2017; Jakarta, Indonesia p. 899. [doi: 10.1109/csci.2017.156]
41. Zarocostas J. How to fight an infodemic. Lancet 2020 Feb;395(10225):676-475. [doi: 10.1016/S0140-6736(20)30461-X]
42. Ministry of Internal Affairs and Communications. Report on Information Circulation Survey on New Coronavirus Infections. URL: https://www.soumu.go.jp/menu_news/s-news/01kiban18_01000082.html [accessed 2022-03-09]

Abbreviations

CAS: Correct Answer Score
CCHL: Communicative and Critical Health Literacy
SNS: social networking service
TV: television
VIF: variance inflation factor
WHO: World Health Organization

Edited by M Gisondi, L Westafer; submitted 29.03.22; peer-reviewed by K Kaphingst, C Mather, L Reid-Chassiakos; comments to author 30.05.22; revised version received 15.06.22; accepted 13.07.22; published 22.07.22.

Please cite as:
Inoue M, Shimoura K, Nagai-Tanima M, Aoyama T
The Relationship Between Information Sources, Health Literacy, and COVID-19 Knowledge in the COVID-19 Infodemic: Cross-sectional Online Study in Japan
J Med Internet Res 2022;24(7):e38332
URL: https://www.jmir.org/2022/7/e38332
doi:10.2196/38332
PMID:35839380

Viewpoint

Use of Twitter Amplifiers by Medical Professionals to Combat Misinformation During the COVID-19 Pandemic

Regina Royan[1], MPH, MD; Tricia Rae Pendergrast[2], BA; Marina Del Rios[3], MS, MD; Shannon M Rotolo[4], PharmD; N Seth Trueger[1,5], MPH, MD; Eve Bloomgarden[6], MD; Deanna Behrens[7], MD; Shikha Jain[8], MD; Vineet M Arora[9], MAPP, MD

[1]Department of Emergency Medicine, Northwestern University, Chicago, IL, United States
[2]Feinberg School of Medicine, Northwestern University, Chicago, IL, United States
[3]Department of Emergency Medicine, Roy J and Lucille A Carver College of Medicine, University of Iowa, Iowa City, IA, United States
[4]Department of Pharmacy, University of Chicago Medicine, Chicago, IL, United States
[5]JAMA Network Open, Chicago, IL, United States
[6]Division of Endocrinology, Northshore University Health System, Evanston, IL, United States
[7]Department of Pediatric Critical Care Medicine, Advocate Children's Hospital, Park Ridge, IL, United States
[8]Division of Hematology and Oncology, University of Illinois, Chicago, IL, United States
[9]Department of Medicine, Pritzker School of Medicine, University of Chicago, Chicago, IL, United States

Corresponding Author:
Regina Royan, MPH, MD
Department of Emergency Medicine
Northwestern University
211 E Ontario St, Suite 200
Chicago, IL, 60611
United States
Phone: 1 312 694 7000
Email: reginaroyan@me.com

Abstract

Social media is an important tool for disseminating accurate medical information and combating misinformation (ie, the spreading of false or inaccurate information) and disinformation (ie, spreading misinformation with the intent to deceive). The prolific rise of inaccurate information during a global pandemic is a pressing public health concern. In response to this phenomenon, health professional amplifiers such as IMPACT (Illinois Medical Professional Action Collaborative Team) have been created as a coordinated response to enhance public communication and advocacy around the COVID-19 pandemic.

(J Med Internet Res 2022;24(7):e38324) doi:10.2196/38324

KEYWORDS

social media; combating disinformation; misinformation; infodemic; amplifier; COVID-19; advocacy; public health communication; disinformation; medical information; health professional amplifier; healthcare profession; health care profession; Twitter; public communication; health information; health promotion

Introduction

In the era of global pandemics, the prolific rise of medical misinformation is a pressing public health concern. The Surgeon General of the United States issued the following directive in 2021: "Health misinformation is a serious threat to public health. It can cause confusion, sow mistrust, harm people's health, and undermine public health efforts. Limiting the spread of health misinformation is a moral and civic imperative that will require a whole-of-society effort" [1]. In early 2019, the American Medical Association issued a letter to the chief executive officers of the country's 6 leading social media and technology companies urging them to ensure their users have access to accurate, timely, and scientifically sound information on vaccines [2]. Despite this call in 2019, many public health professionals found themselves woefully underprepared to face the deluge of misinformation surrounding COVID-19 and the subsequent vaccine rollout [3]. As an organic response to this phenomenon, a new type of professional organization was born: the health professional amplifier.

IMPACT (Illinois Medical Professional Action Collaborative Team) is a 501(c)(3) nonprofit organization designed to help physicians and health professionals engage in grassroots networks, advocate for evidence-based solutions, advise influential stakeholders, and amplify solutions to protect individuals and communities across the state. IMPACT's Twitter account, @IMPACT4HC, is a verified account with 3232 followers.

Twitter has the option to create a private group direct messaging thread, allowing up to 50 individuals to be included in a single messaging conversation. To amplify the message of the @IMPACT4HC account, IMPACT founders (authors VMA, SJ, and EB) created a private group messaging thread of key members across specialties, professions, areas of expertise, and racial and ethnic backgrounds, thus creating an "IMPACT amplifier." The IMPACT amplifier group connects 39 accounts, 7 of which are verified accounts, with over 320,000 followers collectively.

We define an amplifier as a private group of people with a similar mission, who use Twitter for private discussion, brainstorming, and planning, and the subsequent public communication and promotion of the group's work and messaging (Figure 1). This back-channel discussion allows for collaboration between a group of people with similar goals, missions, or agendas, who otherwise may not have open communication channels or perhaps have not even met in person. The goals of an amplifier are to connect people, collaborate, strategize, and amplify ideas on Twitter and other social media platforms.

Since its creation at the start of the pandemic, IMPACT members have used this amplifier to share IMPACT posts and infographics on social media to communicate quickly about pertinent pandemic-related questions and concerns and to stay on top of rapidly changing information across the state. This group of Illinois-based medical professionals, science communication experts, and researchers within IMPACT have used the IMPACT amplifier to facilitate interdisciplinary discussion and coordinate action [4-6]. In this paper, we will discuss this work within the following broad themes: rapid dissemination and promotion of accurate medical information and public health guidance, combating disinformation, and countering harassment.

Figure 1. Rules of engagement for Twitter amplifiers.

Promotion of Timely and Accurate Medical Information

The volume of available health information is both massive and growing; the number of publications in scientific papers is increasing at a rate of over 8% per year, and greater than 1 million new publications are listed in PubMed annually [7]. This amount of information can be overwhelming to both health professionals as well as the lay public and press. The pandemic accelerated the amount of published scientific health information in an unprecedented way. At the same time, the general public needed access to this information rapidly. This deluge of information in the setting of an evolving global pandemic created what the World Health Organization dubbed as an "infodemic" [3]. Scientific claims were being shared without an appreciation of the nuances and limitations of the scientific process, and many incorrect conclusions were shared as fact, in various degrees of good and bad faith [8]. Throughout the pandemic, science communication and public health messaging failed to live up to their potential. The public was often left to decipher confusing and sometimes contradictory information from politicians, agency officials, and an abundance of "armchair epidemiologists," which left many questioning previously trusted sources of public health guidance. The need

for efficient, real-time communication strategies among frontline health care workers, medical professionals, science communicators, and public health experts led to the formation of health communication amplifiers. These amplifiers provided an opportunity to work together to curate content on social media by filtering, interpreting, and amplifying.

First, health professionals can work via amplifiers to select and highlight high-impact information, directing attention to the most relevant studies and data to promote while both implicitly and explicitly prioritizing topics to address and respond to. Next, health professional amplifiers can use their individual and group knowledge, experience, and platform to interpret data, clarify, and explain high-impact research, or to explain why certain findings may not be relevant (eg, clarifying the difference between correlation and causation in observational studies). Importantly, these amplifiers can correct misinformation by calling out incorrect interpretations of research and flawed research methods. Third, health professionals can distill complex information into valuable takeaways (eg, infographics, videos) and distribute them more broadly via an amplifier. Health professionals can also partner with community groups and other local organizations to prevent and address health misinformation [9]. This becomes especially important when engaging with disenfranchised racial and ethnic groups who are often the target of misinformation campaigns or who may not have easy access to culturally relevant and language-concordant reputable sources [10,11]. While many of these objectives can be worked on by individuals, amplifiers can help health professionals find and interpret new data, workshop communication strategies, and share (ie, amplify) and lend credibility to each other's work.

Additionally, many journalists in traditional media use or monitor social media, and amplifiers can help integrate messages and trusted messengers into larger, traditional media opportunities. Health professional amplifiers can also provide a window into the unique perspectives of frontline health care providers. Particularly in the early part of the pandemic, much of the public was sheltered from the suffering and frustrations that health care workers witnessed daily. For example, by sharing the experience of patients saying their last words to loved ones through iPad screens, frontline providers were able to provide a window into the emotional toll of the pandemic. Additionally, the amplifier @IMPACT4HC was able to raise awareness about insufficient personal protective equipment [12].

Lastly, amplifiers can provide a private space for the community, fostering social relationships where professionals can privately share personal and professional news and updates. Having a safe space out of the public discourse to "break plates" with others going through similar professional experiences can be both validating and cathartic. The COVID-19 pandemic, particularly the initial waves, politicization, and public divisiveness, is a unique and often extreme stressor for many health professionals; sharing experiences with peers and colleagues and discovering and discussing the similarities in our experiences was described as therapeutic by a number of our members. Additionally, advocacy on social media is often met with harassment, and the amplifier can likewise serve as a community of colleagues who are going through similar experiences [13]. As an example, one of the coauthors (MDR) was the target of xenophobic and racist attacks after directing remarks in Spanish during a press conference with Illinois government officials highlighting the disproportionate burden of COVID-19 among Latino people. This private space not only provided emotional support but also advice on how to respond to online harassment.

Subspecialty amplifiers created during the pandemic allowed for rapid collaboration and dissemination of specialty-specific medical information, something that was necessary, especially at the start of the pandemic. For example, one of the authors (SJ) created a Women in Medicine amplifier focused on amplifying the work and successes of women across specialties, providing a support network for women in the community who may be harassed online or in person, and a way to share challenges and struggles in a supportive environment. Another author (EB) developed an endocrinology amplifier, which allowed for communication regarding novel presentations of endocrine disorders such as profound hyperglycemia and diabetic ketoacidosis in patients with COVID-19. More broadly, this amplifier allowed endocrinologists to communicate in real time and share information, offer peer support, and amplify one another's work and research. Additionally, when national conferences became virtual, the amplifier was used to promote and support trainee work such as abstracts and oral presentations that otherwise would not have received appropriate notice due to the virtual platform [14].

Combating Misinformation and Disinformation

The widespread dissemination of misleading or false medical and public health information poses a serious threat to health and safety, especially in a pandemic [15]. This can be misinformation (ie, the spreading of false or inaccurate information) or disinformation (ie, spreading misinformation with the intent to deceive) [16]. Social media platforms have become a major source of health information for laypeople [11]. This democratization of health information has numerous benefits and was useful during a pandemic where information rapidly evolved and guidelines were updated as new evidence emerged; however, content related to health and medicine distributed through social media channels is largely unregulated by private platforms. In March 2020 alone, there were 550 million tweets that included the terms "coronavirus," "covid19," or "pandemic," none of which were fact-checked via official mechanisms [17]. Wu and McCormick [15] argue that it is an ethical imperative and professional obligation for physicians to address false or misleading health information on the internet. Many health care professionals feel obligated to undertake efforts to directly combat disinformation, either on a personal level or at the behest of their professional organizations [18,19]. The American Academy of Pediatrics, for example, has provided its members with resources to communicate key messages and combat misinformation [20]; it also provides members with courses and other resources to communicate key messages with families. Because social media is a major platform for

circulating misinformation or disinformation [21], it is also a key place to address or correct it [22].

Amplifiers, such as the IMPACT amplifier on Twitter, help communicate potential messaging of responses (ie, comments, quote tweets) in real time, as well as sharing those messages more widely. They also allow amplifier members a place to discuss among themselves disinformation in need of additional follow-up. For example, within the IMPACT amplifier, a member might share a piece of disinformation that they feel could be better addressed with more nuance in an infographic or a video. This can then be escalated to the team within the coalition developing Myth Buster infographics [23], following the Fact, Myth, Fallacy, Fact refutation strategy for combating misinformation, popularized by climate change activists and described in *The Debunking Handbook* [24]. Other groups such as No License For Disinformation use their own amplifier to facilitate group discussion, identify misinformation, and alert the public and platforms that content is incorrect [25].

Responding to Harassment

A survey study carried out by some of the authors prior to the COVID-19 pandemic found that 1 in 4 physicians reported being personally attacked on social media and 1 in 6 female physicians reported being sexually harassed on social media [13,26]. Online harassment of physicians has likely worsened during the COVID-19 pandemic [27]. In a study published in *Nature*, almost 60% of responding scientists who had commented on COVID-19 to the media or posted on social media said they had experienced attacks on their credibility, and 15% said they had received death threats [28]. Individuals who experience harassment both online and offline report emotional distress and fear. This emotional distress may contribute to moral injury or burnout in physicians who use social media, compounding the effect of the widespread burnout affecting health care professionals during the COVID-19 pandemic [29,30]. Amplifiers may mitigate the professional impact and emotional distress experienced by physicians who are harassed on social media. If a physician belongs to an amplifier, their initial immediate response to harassment can be to share the offending posts with amplifier members. The amplifier provides a semiprivate space for the harassed physician to plan their response or debrief with colleagues. The amplifier can also become a way to protect one another as members of the group are alerted when another member is attacked or targeted by those attempting to spread disinformation. When an individual finds themselves a target, other members of the amplifier respond with data and evidence in the hopes of drowning out bad information with good information, or all members of the amplifier report the offending social media account [27]. For example, when one of our members became the subject of harassment by a radio station account for not working seriously because they were working from home, another member alerted the amplifier members to start responding with stories of burnout, which generated a large number of replies from health care workers about why they were burned out [31].

Future Directions

Our future directions include expanding beyond COVID-19 to other public health topics, such as gun violence and reproductive justice. We recently partnered with AFFIRM, a 501(c)(3) nonprofit organization, to sponsor a Chicago-based discussion on gun violence with a panel of speakers including our chief diversity inclusion officer and cofounder. Our director of community engagement is now crowdsourcing a list of health care workers interested in promoting and advocating for the prevention of gun violence [32]. In addition, we have formalized an internship program with the University of Chicago to have a summer-funded intern to continue to produce infographics to share on social media. Lastly, we have received recognition and a grant from the Association of American Medical Colleges, which provides resources for us to continue to teach trainees how to combat misinformation on social media [33].

Conclusion

Drawing lessons from this experience, it is imperative for medical professionals to utilize all available tools to disseminate accurate medical information and combat disinformation while minimizing harm related to personal and professional harassment that can come with social media advocacy. Understanding the successful use of health professional amplifiers is a substantial way that physicians and other public health professionals can achieve this essential goal.

Conflicts of Interest

NST receives salary support from the American Medical Association as a digital editor for *JAMA Network Open*.

References

1. Murthy VH. Confronting health misinformation: the Surgeon General Advisory on building a healthy information environment. Office of the Surgeon General of the United States. 2021. URL: https://www.hhs.gov/sites/default/files/surgeon-general-misinformation-advisory.pdf [accessed 2022-07-15]
2. AMA urges tech giants to combat vaccine misinformation. American Medical Association. 2019 Mar 13. URL: https://www.ama-assn.org/press-center/press-releases/ama-urges-tech-giants-combat-vaccine-misinformation [accessed 2021-11-23]
3. Managing the COVID-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation. World Health Organization. 2020 Sep 23. URL: https://www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation [accessed 2022-06-10]

4. Hinz G. Lightfoot says Chicagoans should keep wearing masks. Crain's Chicago Business. 2021 May 17. URL: https://www.chicagobusiness.com/greg-hinz-politics/lightfoot-says-chicagoans-should-keep-wearing-masks [accessed 2022-06-10]
5. No ICU beds available in Southern Illinois following rise in COVID hospitalizations. NBC 5 Chicago. 2021 Sep 17. URL: https://www.nbcchicago.com/news/local/no-icu-beds-available-in-southern-illinois-following-rise-in-covid-hospitalizations/2615707/ [accessed 2022-06-10]
6. Fears D, Achenbach J, Martin B. As coronavirus soars, hospitals hope to avoid an agonizing choice: Who gets care and who goes home. The Washington Post. 2020 Nov 11. URL: https://www.washingtonpost.com/climate-environment/2020/11/11/coronavirus-soars-hospitals-hope-avoid-an-agonizing-choice-who-gets-care-who-goes-home/ [accessed 2022-06-10]
7. Landhuis E. Scientific literature: Information overload. Nature 2016 Jul 21;535(7612):457-458. [doi: 10.1038/nj7612-457a] [Medline: 27453968]
8. Patrick M, Venkatesh RD, Stukus DR. Social media and its impact on health care. Ann Allergy Asthma Immunol 2022 Feb;128(2):139-145. [doi: 10.1016/j.anai.2021.09.014] [Medline: 34555532]
9. Rios MD, Giachello A, Khosla S, Luna G, Pobee R, Vergara-Rodriguez P, et al. Illinois Unidos: a community demands equity, justice, and inclusion. Am J Public Health 2021 Oct;111(S3):S204-S207. [doi: 10.2105/AJPH.2021.306407] [Medline: 34709861]
10. Valencia S. Misinformation online is bad in English. But it's far worse in Spanish. The Washington Post. 2021 Oct 28. URL: https://www.washingtonpost.com/outlook/2021/10/28/misinformation-spanish-facebook-social-media/ [accessed 2021-11-29]
11. Khan S, Ramachandran V. Millions depend on private messaging apps to keep in touch. They're ripe with misinformation. PBS. 2021 Nov 05. URL: https://www.pbs.org/newshour/world/millions-depend-on-private-messaging-apps-to-keep-in-touch-theyre-ripe-with-misinformation [accessed 2021-11-29]
12. He S, Ojo A, Beckman AL, Gondi S, Gondi S, Betz M, et al. The story of #GetMePPE and GetUsPPE.org to mobilize health care response to COVID-19: rapidly deploying digital tools for better health care. J Med Internet Res 2020 Jul 20;22(7):e20469 [FREE Full text] [doi: 10.2196/20469] [Medline: 32530813]
13. Pendergrast TR, Jain S, Trueger NS, Gottlieb M, Woitowich NC, Arora VM. Prevalence of personal attacks and sexual harassment of physicians on social media. JAMA Intern Med 2021 Apr 01;181(4):550-552 [FREE Full text] [doi: 10.1001/jamainternmed.2020.7235] [Medline: 33393965]
14. Monostra M. Endocrinologists turn to social media to connect with public, colleagues during pandemic. Healio. 2021 Apr 21. URL: https://www.healio.com/news/endocrinology/20210415/endocrinologists-turn-to-social-media-to-connect-with-public-colleagues-during-pandemic [accessed 2021-11-27]
15. Wu JT, 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]
16. Gebel M. Misinformation vs. disinformation: What to know about each form of false information, and how to spot them online. Business Insider. 2021 Jan 15. URL: https://www.businessinsider.com/misinformation-vs-disinformation [accessed 2022-06-10]
17. Understanding the infodemic and misinformation in the fight against COVID-19. Pan American Health Organization. 2020. URL: https://iris.paho.org/bitstream/handle/10665.2/52052/Factsheet-infodemic_eng.pdf [accessed 2022-07-15]
18. Bautista JR, Zhang Y, Gwizdka J. US physicians' and nurses' motivations, barriers, and recommendations for correcting health misinformation on social media: qualitative interview study. JMIR Public Health Surveill 2021 Sep 01;7(9):e27715 [FREE Full text] [doi: 10.2196/27715] [Medline: 34468331]
19. Arora VM, Bloomgarden E, Jain S. Supporting health care workers to address misinformation on social media. N Engl J Med 2022 May 05;386(18):1683-1685. [doi: 10.1056/NEJMp2117180] [Medline: 35353977]
20. Jenco M. AAP, White House encourage clinicians to combat COVID vaccine misinformation. AAP News. 2021 Nov 18. URL: https://publications.aap.org/aapnews/news/18771/AAP-White-House-encourage-clinicians-to-combat?autologincheck=redirected [accessed 2021-11-29]
21. Darius P, Urquhart M. Disinformed social movements: a large-scale mapping of conspiracy narratives as online harms during the COVID-19 pandemic. Online Soc Netw Media 2021 Nov;26:100174 [FREE Full text] [doi: 10.1016/j.osnem.2021.100174] [Medline: 34642647]
22. Walter N, Brooks JJ, Saucier CJ, Suresh S. Evaluating the impact of attempts to correct health misinformation on social media: a meta-analysis. Health Commun 2021 Nov 06;36(13):1776-1784. [doi: 10.1080/10410236.2020.1794553] [Medline: 32762260]
23. Rotolo SM, Jain S, Dhaon S, Dokhanchi JK, Kalata E, Shah T, et al. A coordinated strategy to develop and distribute infographics addressing COVID-19 vaccine hesitancy and misinformation. J Am Pharm Assoc (2003) 2022;62(1):224-231 [FREE Full text] [doi: 10.1016/j.japh.2021.08.016] [Medline: 34483058]
24. Lewandowsky S, Cook J, Ecker U, Albarracin D, Amazeen M, Kendou P, et al. The Debunking Handbook 2020. 2020. URL: https://www.climatechangecommunication.org/wp-content/uploads/2020/10/DebunkingHandbook2020.pdf [accessed 2022-07-15]
25. Rubin R. When physicians spread unscientific information about COVID-19. JAMA 2022 Mar 08;327(10):904-906. [doi: 10.1001/jama.2022.1083] [Medline: 35171224]

26. Arora V, Pendergrast T, Jain S. Doctors on social media are being cyberbullied and sexually harassed. It's time to change that. STAT. 2021 Jan 05. URL: https://www.statnews.com/2021/01/05/doctors-on-social-media-are-being-cyberbullied-and-sexually-harassed-its-time-to-change-that/ [accessed 2021-11-29]
27. Larkin H. Navigating attacks against health care workers in the COVID-19 era. JAMA 2021 May 11;325(18):1822-1824. [doi: 10.1001/jama.2021.2701] [Medline: 33881489]
28. Nogrady B. 'I hope you die': how the COVID pandemic unleashed attacks on scientists. Nature 2021 Oct;598(7880):250-253. [doi: 10.1038/d41586-021-02741-x] [Medline: 34645996]
29. Berg S. Half of health workers report burnout amid COVID-19. American Medical Association. 2021 Jul 20. URL: https://www.ama-assn.org/practice-management/physician-health/half-health-workers-report-burnout-amid-covid-19 [accessed 2022-03-27]
30. Standard definitions: final dispositions of case codes and outcome rates for surveys. American Association for Public Opinion Research. 2016. URL: https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf [accessed 2022-07-15]
31. Bowen A. 'We're getting attacked.' Even before COVID-19, doctors faced online harassment, including death threats. The pandemic makes it worse. Chicago Tribune. 2021 Jan 06. URL: https://www.chicagotribune.com/coronavirus/ct-life-covid-health-care-workers-harassment-pandemic-tt-20210106-u5praqcvuveptp3xi2n2yo4wf4-story.html [accessed 2022-06-10]
32. Reframe Chicago. AFFIRM, The Aspen Institute. 2020 Oct 21. URL: https://affirmresearch.org/past-events/reframe-chicago [accessed 2022-06-10]
33. 2022 AAMC Innovations That Bolster Community Trust in Science Award. Association of American Medical Colleges. URL: https://www.aamc.org/what-we-do/aamc-awards/innovations-in-research [accessed 2022-06-10]

Abbreviations

IMPACT: Illinois Medical Professional Action Collaborative Team

Edited by M Gisondi, J Faust; submitted 13.04.22; peer-reviewed by M Lotto, JR Bautista; comments to author 08.05.22; revised version received 10.06.22; accepted 13.07.22; published 22.07.22.

Please cite as:
Royan R, Pendergrast TR, Del Rios M, Rotolo SM, Trueger NS, Bloomgarden E, Behrens D, Jain S, Arora VM
Use of Twitter Amplifiers by Medical Professionals to Combat Misinformation During the COVID-19 Pandemic
J Med Internet Res 2022;24(7):e38324
URL: https://www.jmir.org/2022/7/e38324
doi:10.2196/38324
PMID:35839387

Original Paper

COVID-19 Misinformation and Social Network Crowdfunding: Cross-sectional Study of Alternative Treatments and Antivaccine Mandates

Nathan M Shaw[1], MD; Nizar Hakam[1], MD; Jason Lui[1], BSc; Behzad Abbasi[1], MD; Architha Sudhakar[1], MD; Michael S Leapman[2], MD; Benjamin N Breyer[1,3], MD, MSc

[1]Department of Urology, University of California, San Francisco, San Francisco, CA, United States

[2]Department of Urology, Yale University School of Medicine, New Haven, CT, United States

[3]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:
Nathan M Shaw, MD
Department of Urology
University of California, San Francisco
1001 Potrero Ave
Suite #3A26
San Francisco, CA, 94110
United States
Phone: 1 9252859854
Email: shawlah88@gmail.com

Abstract

Background: Crowdfunding is increasingly used to offset the financial burdens of illness and health care. In the era of the COVID-19 pandemic and associated infodemic, the role of crowdfunding to support controversial COVID-19 stances is unknown.

Objective: We sought to examine COVID-19–related crowdfunding focusing on the funding of alternative treatments not endorsed by major medical entities, including campaigns with an explicit antivaccine, antimask, or antihealth care stances.

Methods: We performed a cross-sectional analysis of GoFundMe campaigns for individuals requesting donations for COVID-19 relief. Campaigns were identified by key word and manual review to categorize campaigns into "Traditional treatments," "Alternative treatments," "Business-related," "Mandate," "First Response," and "General." For each campaign, we extracted basic narrative, engagement, and financial variables. Among those that were manually reviewed, the additional variables of "mandate type," "mandate stance," and presence of COVID-19 misinformation within the campaign narrative were also included. COVID-19 misinformation was defined as "false or misleading statements," where cited evidence could be provided to refute the claim. Descriptive statistics were used to characterize the study cohort.

Results: A total of 30,368 campaigns met the criteria for final analysis. After manual review, we identified 53 campaigns (0.17%) seeking funding for alternative medical treatment for COVID-19, including popularized treatments such as ivermectin (n=14, 26%), hydroxychloroquine (n=6, 11%), and vitamin D (n=4, 7.5%). Moreover, 23 (43%) of the 53 campaigns seeking support for alternative treatments contained COVID-19 misinformation. There were 80 campaigns that opposed mandating masks or vaccination, 48 (60%) of which contained COVID-19 misinformation. Alternative treatment campaigns had a lower median amount raised (US $1135) compared to traditional (US $2828) treatments ($P$<.001) and a lower median percentile of target achieved (11.9% vs 31.1%; P=.003). Campaigns for alternative treatments raised substantially lower amounts (US $115,000 vs US $52,715,000, respectively) and lower proportions of fundraising goals (2.1% vs 12.5%) for alternative versus conventional campaigns. The median goal for campaigns was significantly higher (US $25,000 vs US $10,000) for campaigns opposing mask or vaccine mandates relative to those in support of upholding mandates (P=.04). Campaigns seeking funding to lift mandates on health care workers reached US $622 (0.15%) out of a US $410,000 goal.

Conclusions: A small minority of web-based crowdfunding campaigns for COVID-19 were directed at unproven COVID-19 treatments and support for campaigns aimed against masking or vaccine mandates. Approximately half (71/133, 53%) of these campaigns contained verifiably false or misleading information and had limited fundraising success.

International Registered Report Identifier (IRRID): RR2-10.1001/jamainternmed.2019.3330

(J Med Internet Res 2022;24(7):e38395) doi:10.2196/38395

KEYWORDS

COVID-19; misinformation; infodemic; social media; GoFundMe; vaccine hesitancy; vaccination; infodemiology; campaign; treatment; vaccine mandate; health care; online health information

Introduction

The COVID-19 pandemic has resulted in tremendous health and financial consequences worldwide, with an estimated loss of 24 trillion dollars in the first year of the pandemic alone [1]. During the pandemic, many turned to crowdfunding to cover health expenditures, as they lost business revenue and were confronted with other new financial burdens; this increase has been similarly seen in the use of crowdfunding for other illnesses [2,3]. The initial increase in funding requests—150,000 campaigns (not all COVID-19 related) in March 2020 alone—were so dramatic that it prompted public comments to the US Congress from the GoFundMe website [4]. GoFundMe remains the largest crowdfunding service, and prior to the pandemic over one-third of the campaigns were used for medical purposes [5-8].

The pervasiveness of unproven or disproven theories about COVID-19—particularly the safety and efficacy of vaccination, nonpharmaceutical interventions such as masking, and potential treatments—poses a major and persistent public health challenge. The "infodemic"—defined as an "overwhelming, complex, and contradictory information... including fake news... about the origins of the virus... treatment options unsupported by data... and the life-saving vaccine"—has accompanied the COVID-19 pandemic [9]. The volume and pace of new information related to vaccination, masking, health care burnout, compassion fatigue, novel variants, and alternative treatments have changed the landscape and discussion around COVID-19 [10-12]. Furthermore, data continue to suggest that misleading COVID-19 information remains prevalent across search and social media platforms [13-18]. One study examined the use of GoFundMe to raise money for unproven COVID-19 prophylactic medications and suggested that GoFundMe be included as part of the conversation around COVID-19 misinformation [19]. As the world enters the third year of the pandemic, there have been media reports that GoFundMe has continued to be used to fundraise for COVID-19 misinformation campaigns, occasionally resulting in removal by GoFundMe [20,21]. While there is no stated policy on GoFundMe regarding COVID-19 misinformation, a press release noted that "Fundraisers raising money to promote misinformation about vaccines violate GoFundMe's terms of service and will be removed from the platform" [22]. There are no recent data to examine the use of the platform as a means to disseminate unproven, disproven, or misleading COVID-19 information and to fundraise for these causes.

We conducted a comparative analysis of crowdfunding campaigns for COVID-19–related requests. We analyzed campaigns for financial outcomes to better understand the economic needs of this population and the ongoing social appetite for crowdfunding COVID-19–related issues. We specifically sought to evaluate fundraising seeking assistance for unproven, disproven, or misleading information, including alternative treatments not endorsed by public health agencies. In addition, we examined fundraising aimed at opposition to COVID-19 mitigation efforts, including masking and vaccination. Finally, we examined the prevalence of COVID-19 misinformation among these campaigns.

Methods

Data Source and Study Design

We performed a cross-sectional analysis of GoFundMe campaigns for individuals requesting donations for COVID-19 relief. All campaigns that were open or collected donations on GoFundMe from September 2021 to November 2021 were screened. Campaigns in which the story, title, or category included COVID-19 or the following key words were included: "COVID-19," "Corona Virus," "China Virus," "SARS-CoV2," "Pandemic," "Right to Try*," "Vit* D*," "Hydroxychloroquine," "Ivermectin," "FrontLine COVID-19 Critical Care Alliance (FLCCC)," "Vaccine*," "Mask*," "Ventilator," "Hospital*," "Nitrous oxide (NOS)," "Melatonin," "CoronaBox," "Essential Oils," "Herbal*," "Homeopath*," "Complimentary," and "*Mandate*." Key words were generated based on publicly available lists of frequently searched COVID-19–related "unproven" treatments and myths (Centers for Disease Control and Prevention, World Health Organization, Wikipedia, WebMD, Mayo Clinic, and Johns Hopkins) [23-27]. The list was reviewed and finalized in consultation with an expert author (ML). Asterix (*) indicates a variable to incorporate multiple suffixes or prefixes. For example, Homeopath* would collect searches in which Homeopathic or Homeopathy were included.

These key words were queried on the GoFundMe platform across 50 US states for 1200 searches (50 states x 24 key words = 1200 searches). Of note, the use of states improves search algorithm by limiting the individual search results in Python (Python Software Foundation) without changing the campaigns that were captured [8]. A custom Python programming language code was used to automatically retrieve information from publicly available campaign webpages. This search was conducted in November 2021. Manual review took place in February of 2022. This data collection methodology has been previously described [28,29].

Categories and Manual Review

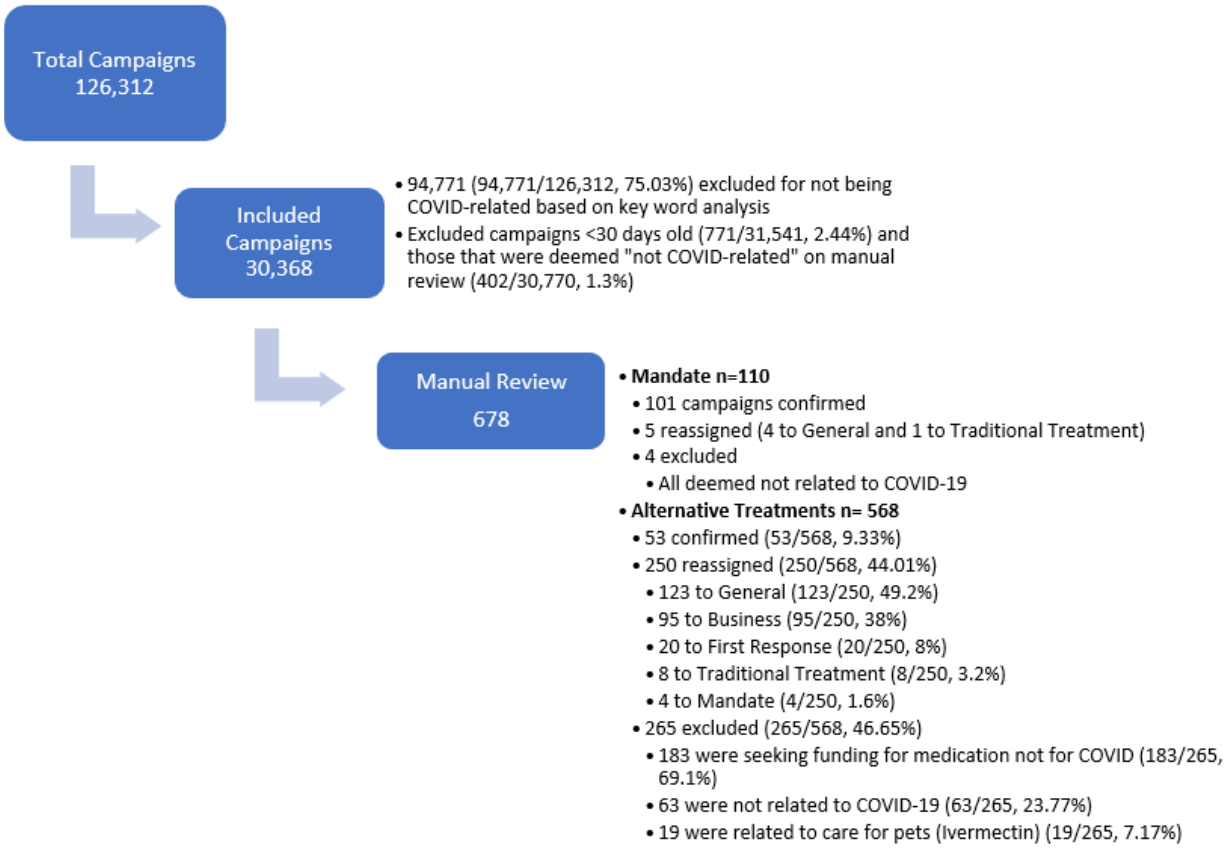
The initial study population was filtered by the same key words to ensure the inclusion of "COVID-19," "Pandemic," "Corona Virus," "China Virus," or "SARS-CoV2" in the title or story of the campaign in relation to the pandemic or its effect. Further key word categorization was performed to divide the selection into categories that were chosen a priori, such as "Alternative treatments," "Business-related," "First Response," "Mandate," "Traditional Treatments," and "General" (Multimedia Appendix

1). “Alternative treatments” was defined by key words within the campaign (Multimedia Appendix 1), which focused on the use of unproven or complementary therapies for COVID-19 (eg, “The specific goal of this fundraiser is to support [treatment] with Ivermectin”). Moreover, “Business-related” was defined by key words (Multimedia Appendix 1) that focused on funding for lost income or commercial revenue (eg, “Covid was devastating for my businesses and I wasn't eligible for jobkeeper so I've been on a bit of a knife edge for the last year”). “First Response” was defined by key words (Multimedia Appendix 1) that focused on funding for COVID-19 first responders, mission trips, or community organizations (eg, “Unfortunately, these hardworking men and women [nurses] don't have the necessary protection they need to stay safe as they save lives so that they can continue to save lives”). “Mandate” was defined by key words (Multimedia Appendix 1) that focused on funding for legal relief from COVID-19–related restrictions (eg, “Breathe Free Colorado was founded to stop the unconstitutional mask mandate created by Governor through executive order”). “Traditional Treatments” was defined by key words (Multimedia Appendix 1) that focused on funding for COVID-19 hospital or other health expenditures (eg, “Mickey has contracted covid19 and is in the UIHC in Iowa City on a ventilator and very sick. Mickey was the household income so they're needing help plus hospital Bill's are piling up. Please help if you can”). All other campaigns that did not fall into the established categories based on key word were defined as “General.” A hierarchy was established such that campaigns in which multiple key words were present were preferentially assigned to (1) “Alternative Treatment,” (2) “Mandate,” (3) “Traditional Treatments,” (4) First Response,” and (5) “General.” This hierarchy was chosen a priori, knowing categories 1 and 2 would be manually reviewed.

All campaigns identified by key word in the categories “Mandate” and “Alternative Treatment” were manually reviewed. Manual review confirmed that the campaign was seeking funding relief for the alternative treatment or mandate-related expenses and confirmed the web-scraped variables (eg, funding goal and total funds raised) were accurate and up-to-date. Furthermore, manual review of the “Mandate” category facilitated further analysis of whether the mandate-related campaigns were pro- or antimask and pro or antivaccine. Manual review was also performed for a random sample of 100 campaigns from each other category (“Business-related,” “First Response,” “Mandate,” “Traditional Treatments,” and “General”) as a similar sensitivity analysis. Campaigns were permitted to be moved from one category to another or excluded on manual review.

Manual review was also undertaken to capture campaigns containing explicit COVID-19 misinformation. COVID-19 misinformation for the purposes of this study adapted the standard definition of “false or misleading information meant to deceive” [30] to include only those claims where direct cited evidence could be provided to refute a claim. For each misleading statement, a reference is provided to directly refute this claim. Vague claims or those that could not be possible to prove false were not considered misinformation. For example, a campaign claiming individual side effects from the vaccine, which could not possibly be verified, would not be coded as containing misinformation. Similarly, charged language around vaccination (eg, “jab of Satan”) was not considered misinformation. Statements such as “masks cause harm” or “children cannot get COVID-19” were considered misinformation, and references were provided demonstrating the claims as false. Finally, the manual review stated if the campaign was withdrawn. The full review process is summarized in Figure 1. The design and reporting of this study adhered to the guidelines of the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement (Multimedia Appendix 2) [31].

Figure 1. Inclusion of COVID-19–related campaigns.



Variables

For each campaign, basic narrative features, engagement metrics, and financial variables were extracted. We examined campaign titles, narratives, creation date, number of social media shares, campaign goal amount, number of donations, and amount of funding raised. Among those that were manually reviewed, the variables of "Mandate Type," "Prorestriction," and "Antirestriction" were also included. Campaigns were also reviewed for COVID-19 misinformation. As described above, this adhered to the strict definition of verifiably false information included in the title, description, or story of the included campaign. Within the "Alternative Treatment" category, the individual drug or treatments were reviewed and included in the analysis.

Statistical Analysis

We used descriptive statistics to characterize the study cohort. Continuous variables skewed distribution and were reported as median (interquartile range). Campaign success was defined as those that received funding amounts equal to the declared goal or higher. Campaigns with no funding goal were considered neither successful nor unsuccessful. Campaigns that were not open for at least 30 days at the time of data collection were excluded. We reported the proportion of campaigns deemed successful as well as the median percentile of funding goal received. Characteristics of successful and unsuccessful campaigns were compared using the chi-square or Fisher exact tests for categorical variables (campaign success, median goal, median raised, median donors, median donations, and campaign type) or Mann Whitney *U* test for continuous variables (donors, donations, total amount raised, total campaign goal, photos, updates, comments, shares, and followers). In a similar fashion, we compared campaigns in the "Traditional Treatments" category to those in the "Alternative Treatments" category, and we compared those supporting vaccine or mask mandates to those opposing vaccine or mask mandates within the "Mandates" category. We used Stata 17 (StataCorp), and we considered a 2-sided *P* value of <.05 as statistically significant.

Ethics Approval

The data collected from GoFundMe were deidentified and did not qualify under the University of California Institutional Review Board (IRB) as human subjects research. Therefore, IRB approval was not sought, and appropriate IRB self-certification was completed and maintained per University of California San Francisco policy.

Results

Campaigns

A total of 126,695 campaigns were identified for analysis, 126,312 (99.7%) of which were active from April 2011 to November 2021. Of those, 30,368 (24.0%) campaigns met the criteria for final analysis (Figure 1). Moreover, 6444 campaigns were seeking aid for medical costs associated with COVID-19. The majority of these (6391/6444, 91.2%) were seeking aid for traditional treatments including hospital or intensive care unit

care (4285/6444, 66.5%), medical bills or hospital bills (1569/6444, 24.3%), remdesivir (63/6444, 0.98%), monoclonal antibody treatment (16/6444, 0.25%), and home oxygen (13/6444, 0.20%; Multimedia Appendix 1). Based on key word, there were 568 campaigns assigned to alternative treatments. After excluding those classified as "General," "Business," "First Response," Traditional Treatment," and "Mandate," there were 53 campaigns (53/30,368, 0.17%) that were expressly seeking funding for unproven or alternative treatments for COVID-19 (Table 1). These campaigns requested donations including stories asking for herbal remedies (17/53, 32%), ivermectin (14/53, 26%), hydroxychloroquine (6/53, 11%), and vitamin D (4/53, 7.5%).

Table 1. Campaign characteristics and predictors of campaign success.

Category	Values			
	All	Unsuccessful[a]	Successful	*P* value[b]
Total, n (%)	30,368 (100)	25,502 (86.9)	3847 (13.1)	.003
Alternative treatments, n (%)	53 (0.17)	44 (88)	6 (12)	—[c]
Mandate, n (%)	96 (0.32)	90 (95.7)	4 (4.3)	—
First response, n (%)	3260 (10.7)	2691 (86.5)	421 (13.5)	—
Traditional treatments, n (%)	6391 (21)	5270 (85.7)	883 (14.4)	—
Businesses, n (%)	415 (1.4)	353 (87.4)	51 (12.6)	—
General, n (%)	20,153 (66.4)	17,054 (87.3)	2482 (12.7)	—
Number of followers, median (IQR)	27 (4-75)	21 (2-62)	81 (39-172)	<.001
Number of donations, median (IQR)	21 (2-58)	16 (1-48)	63 (30-134)	<.001
Number of donors, median (IQR)	20 (2-55)	15 (1-46)	60 (29-127)	<.001
Ratio donations:donors, median (IQR)	1.01 (1-1.05)	1.006 (1-1.05)	1.019 (1-1.05)	<.001

[a]Some campaigns did not have funding goal despite remaining open. This accounts for the small number missing between the total n and those included in the campaign success columns.

[b]Statistical tests comparing successful vs unsuccessful campaigns.

[c]Not available or appropriate.

Alternative Versus Traditional Treatment

Among the 53 campaigns identified as seeking funds for unproven COVID-19 treatment, the median fundraising goal (US $10,750) was higher than campaigns seeking money for traditional treatment (US $10,000), and fewer alternative campaigns met their stated goal (9/53, 16.9%) compared to traditional (1121/6391, 17.5%; *P*=.92). Alternative treatment campaigns had a significantly lower median amount raised compared with traditional treatments (US $1135 vs US $2828, respectively; *P*<.001) and a lower median percentile of target achieved at 11.9% for alternative compared to 31.1% for traditional treatments (*P*=.003). Taken as an aggregate, traditional treatments raised 12.5% of the total requested funding goal while alternative treatments reached only 2.1% (Table 2; Figure 2). On manual review, the 6 (6/53, 11%) campaigns in the alternative treatment that met their goal were substantively different from the unsuccessful campaigns. Moreover, 2 (2/53, 4%) campaigns were written for family members who were severely ill in the intensive care unit requesting funds to add alternative treatments to traditional as "last resorts." An additional 2 (2/53, 4%) campaigns were seeking funding for herbal remedies to help whole communities—Navajo nation and an unspecified "ancient folk" community. Only 2 (2/53, 4%) successful campaigns were written by the individual seeking funds, and both cited multiple other hardships and medical comorbidities in addition to asking for funding for alternative treatment (ivermectin in both cases). Another successful campaign was seeking funding for a study to evaluate the efficacy of vitamin D for treatment and prevention of COVID-19. Interestingly, despite its success, this campaign was withdrawn during the period of manual review. Finally, the highest grossing successful and unsuccessful alternative treatment campaigns (US $12,555 donated from US $100 asked and US $11,684 donated from US $30,000 asked) were seeking funding for frontline medical personnel, including the use of vitamin D and ivermectin for the treatment and prevention of COVID-19.

Alternative treatments contained COVID-19 misinformation in 23 (43%) campaigns including advocacy on alternative treatments over vaccination or traditional treatments (n=18, 34%), claims that the pandemic is not real or is a conspiracy (n=2, 4%), and false claims about the safety or efficacy of the vaccine (n=3, 7%). From the manual review of the subset (n=100) of traditional treatments, none of them contained COVID-19 misinformation. A total of 5 (5/53, 9.4%) alternative treatment campaigns were withdrawn by the manual review period compared to 1 (1/100, 1%) in the traditional group (Table 3).

Table 2. Comparison of traditional and alternative treatment campaigns.

Characteristics	Alternative treatments	Traditional treatments	*P* value
Total campaigns	53	6391	—[a]
Success, n (%)	9 (16.9)	1121 (17.5)	.92
Total amount raised (US $)	115,463	52,715,762	—
Total amount goal (US $)	5,600,725	421,223,293	—
Total achieved (%)	2.06	12.5	—
Campaign features			
Goal (US $), median (IQR)	10,750 (2700-30,000)	10,000 (5000-23,000)	.81
Funds raised (US $), median (IQR)	1135 (50-2975)	2828 (665-7788)	<.001
Funds raised[b] (%), median (IQR)	11.9 (0.01-38.9)	31.1 (6.8-74.9)	.03
Photos, median (IQR)	1 (1-3)	1 (1-2)	.76
Number of updates, median (IQR)	0 (0-3)	1 (0-2)	.78
Number of comments, median (IQR)	0 (0-2)	1 (0-3)	.009
Number of donations, median (IQR)	17 (1-36.5)	35 (10-85)	.001
Number of donors, median (IQR)	16.5 (1-32)	33 (9-81)	.001
Ratio donations:donors, median (IQR)	1.017 (1-1.09)	1.014 (1-1.047)	.5
Number of shares, median (IQR)	26.5 (0-219)	113 (5-398)	.007
Number of followers, median (IQR)	22 (1-47)	45 (13-109)	.001

[a]Analyses not performed or not appropriate.

[b]Total funds raised divided by requested. This only applied to campaigns with a funding goal.

Figure 2. Percent of total funding goal reached in US dollars.

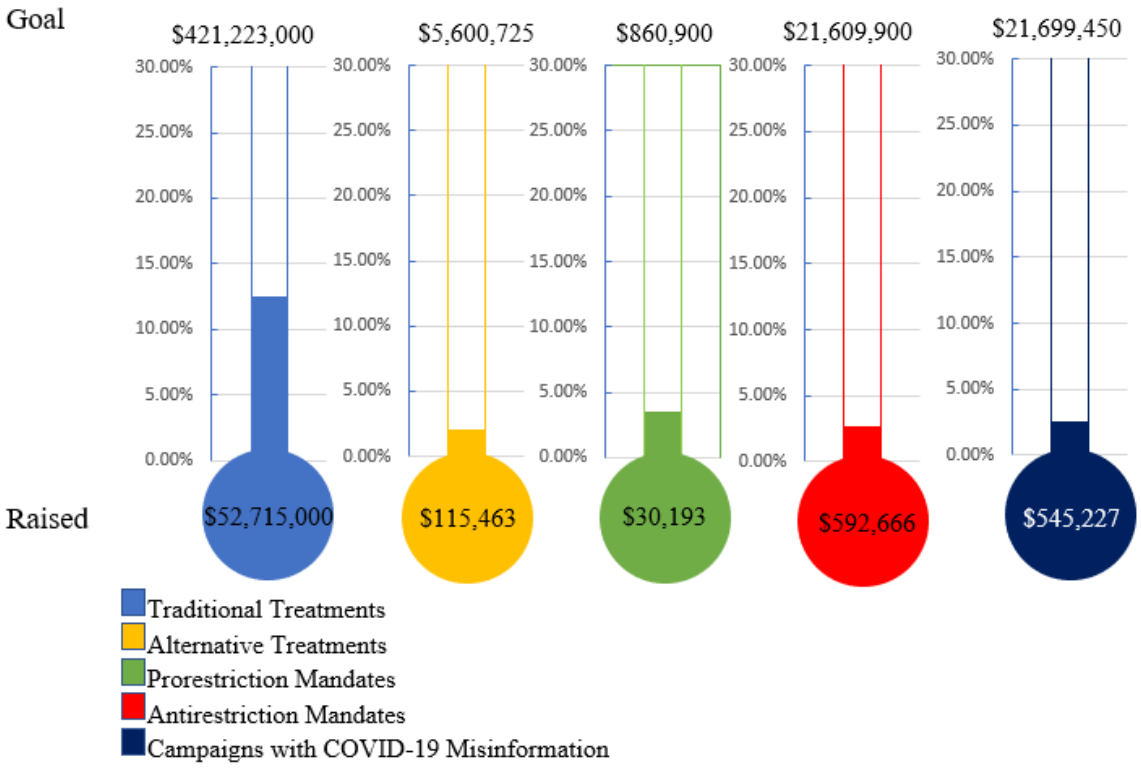


Table 3. Summary of unproven "alternative treatment" campaigns.

Alternative treatments	All, n (%)	Successful n (%)	Funds raised (US $), n (% of requested)	Funds requested (US $)	COVID-19 misinformation, n (%)	Withdrawn, n (%)
Total	53 (100)	7 (13.2)	115,463 (2.06)	5,600,725	23 (43)	5 (9.4)
Herbal or homeopathic remedies	17 (32)	3 (17.6)	20,410 (0.44)	4,606,475	7 (41)	2 (3.7)
Ivermectin	14 (26)	1 (7.1)	25,615 (10.7)	238,800	11 (79)	1 (1.8)
Hydroxychloroquine	6 (11)	2 (33.3)	42,728 (17.5)	242,100	0 (0)	1 (1.8)
Vitamin D	4 (7.5)	0 (0)	4478 (12.4)	36,000	3 (75)	1 (1.8)
Meditation or functional	2 (4)	0 (0)	2063 (0.81)	252,700	0 (0)	0 (0)
Naturopathic	2 (4)	0 (0)	11784 (23.4)	50,250	1 (50)	0 (0)
Air purification	1 (2)	0 (0)	130 (0.13)	100,000	0 (0)	0 (0)
Essential oils	1 (2)	0 (0)	1135 (9.9)	11,500	0 (0)	0 (0)
Lifestyle	1 (2)	0 (0)	0 (0)	15,000	0 (0)	0 (0)
Music	1 (2)	0 (0)	0 (0)	30,000	0 (0)	0 (0)
Native American remedies	1 (2)	1 (100)	918 (153)	600	0 (0)	0 (0)
Melatonin	1 (2)	0 (0)	1880 (15.6)	12,000	0 (0)	0 (0)
Soaps	1 (2)	0 (0)	1197 (66.5)	1800	0 (0)	0 (0)
Vitamin C	1 (2)	0 (0)	3125 (90)	3500	1 (100)	0 (0)

Mandates

A total of 96/30,368 (0.32%) campaigns were created to fund political positions in response to mask or vaccine mandates. Of these 93 campaigns, 80 (83%) were created to oppose restrictions (against requiring masks or vaccines; the most frequent mask mandates were in schools). The vast majority of campaigns did not meet funding goal (90/96, 86.9%), and the median funding percentage (proportion of goal reached) was 14.2%.

The median goal for campaigns was significantly higher (US $25,000 vs US $10,000) for campaigns opposing mask or vaccine mandates (P=.04). Almost all (14/16, 87.5%) campaigns supporting restrictions (prorestriction) were in support of reinstating mask mandates, particularly in schools in areas where the mandate had been or was going to be lifted (Table 4). Category, followers, donations, and ratio of donations:donors were significantly associated with campaign success (Table 1). Among the campaigns seeking changes to health care mandates, one campaign seeking broad legal challenges against a state governor obtained US $31,730 (42.3%) of the US $75,000 requested. The remaining 6 campaigns obtained US $622 (0.2%) out of US $410,000. This is in contrast with campaigns seeking to re-enforce masking in schools which raised $37,247 (4.9%) of $761,400.

Among the 80 campaigns requesting support for antirestriction positions, 48 (60%) contained verifiably false claims regarding COVID-19 and COVID-19 vaccination. This included claims that masks were ineffective or dangerous (n=18, 22.5%), children are unaffected by COVID-19 (n=10, 12.5%), COVID-19 is mild or similar to the common cold (n=10, 12.5%), individuals would be better to seek nonvaccine prevention (n=6, 7.5%), equivocal efficacy between various unproven COVID-19 treatments (n=4, 5%), the pandemic is not real or is a conspiracy (n=2, 2.5%), and the vaccine is not safe or effective (n=1, 1%; Table 5). Comparatively, the small minority of campaigns seeking to reimpose restrictions (16) contained no COVID-19 misinformation.

Table 4. Mandate-related campaigns.

Characteristics of the campaigns	Prorestriction[a] (n=16)	Antirestriction[b] (n=80)	*P* value
Median goal in US $, n (IQR)	10,000 (2000-25,000)	25,000 (10,000-75,000)	.04
Median raised in US $, n (IQR)	283 (0-3120)	210 (0-5131)	.82
Successful campaigns, n/N (%)[c]	2/15 (13.3)	2/79 (2.5)	.12
Number of donations, median (IQR)	4 (0-34)	5 (0-83.5)	.65
Number of donors, median (IQR)	4 (0-33)	5 (0-79.5)	.66
Mandate type, n (%)			.007
Mask	14 (87.5)	35 (43.8)	
Vaccine	2 (12.5)	37 (46.3)	
All restrictions	0 (0)	8 (10)	
Mandate setting, n (%)			.01
School	11 (68.8)	32 (40)	
Job	0 (0)	22 (27.5)	
Health care	0 (0)	7 (8.8)	
None or general	5 (31.3)	19 (23.8)	
Contains COVID-19 misinformation, n (%)	0 (0)	48 (60)	<.001
Funds (US $) received of total requested, n/N (%)			—[d]
Total	30,193/860,900 (3.51)	514,584/21,419,900 (2.4)	
Mask	29,626/833,400 (3.55)	403,964/11,583,000 (3.49)	
Vaccine	567/27,500 (2.06)	110,620/9,836,900 (1.12)	

[a]Prorestriction indicates campaigns that were seeking to reinstate, strengthen, or support existing mandates on masking or vaccination. A recurring example was parents seeking support to reinstate mask mandates in schools that had lifted mask mandates.

[b]Antirestriction indicates campaigns that were seeking to remove mandates that required masking or vaccination.

[c]Some campaigns did not have funding goal despite remaining open. This accounts for the small number missing between the total n and those included in the campaign success columns.

[d]Not available or appropriate.

Table 5. Description of COVID-19 misinformation in GoFundMe campaigns.

Misinformation	Value (N=71[a]), n (%)	True statement with reference
[Alternative Treatment] is equivalent to traditional treatments or can replace vaccination.	22 (31)	"There is no evidence to support the efficacy of alternative treatments. There is decreased COVID-19 transmission and hospitalization rate among vaccinated individuals" [32].
Masks are ineffective or dangerous.	18 (25)	Masks are effective in decreasing the spread of COVID-19 [33].
Children cannot get COVID-19 or are not seriously affected.	10 (14)	"Children are less likely to experience severe COVID-19 compared to adults, but still have concerning rates of hospitalization and serious illness" [34].
COVID-19 is not severe or is equivalent to the common cold.	10 (14)	"There is a clinical spectrum including severe illness and death resulting from COVID-19 infection" [35].
Government is forcing an unsafe or experimental vaccine	6 (8)	"Mandates are in place for certain professions" [36]. "Data on vaccine safety is robust" [37].
Vaccination results in severe side effects.	4 (6)	"Vaccine side effects are rare and overwhelmingly mild" [37,38].
The pandemic is fake or a conspiracy.	4 (6)	"There are numerous and changing conspiracies about the origins and reality of the COVID-19 pandemic" [39].

[a]Some campaigns contained more than one verifiably false statement, so the sum of categories is greater than total campaigns containing false statements.

Discussion

Principal Results

There is a minority on the social media site GoFundMe seeking financial support for fringe beliefs regarding COVID-19. Among the small number of campaigns seeking assistance for alternative treatments and to oppose mandates, the majority (71/133, 53%) contained verified false information about COVID-19. These findings are novel in the literature for crowdfunding and verify the trends occurring amid the infodemic across other social media sites and the popular reporting on GoFundMe [20,21,40-42].

A small minority of web-based crowdfunding campaigns for COVID-19 on the GoFundMe platform explicitly sought funding for unproven or disproven medical treatments, as well as opposition to public health interventions such as masking (Table 5). Crowdfunding campaigns for alternative COVID-19 treatments and antimasking generated fewer views and less funding compared to campaigns seeking funding for traditional medical treatment. Interestingly, the number of campaigns that reached their stated goal was similar between alternative treatments and traditional treatments. Despite the similarity in "successful campaigns," the funds generated as a percentage of the total requested between alternative (US $115,000/$5,600,000, 2.06%) and traditional treatments (US $52,715,000/$421,223,000, 12.5%) were substantially different. The discrepancy indicates that a small number of campaigns are able to generate significant funding while the majority fail to generate even a small percentage of the requested funds. These findings highlight the pervasiveness of unproven or disproven information about COVID-19, and the complex media through which this information continues to propagate.

This study is the first to examine the use of crowdfunding for positions on COVID-19–related mandates. We found that campaigns seeking funding for positions on mask and vaccine mandates rarely met funding goals. The majority of campaigns were in opposition to COVID-19 restrictions. In total, 48/80 (60%) of the antirestriction campaigns contained COVID-19 misinformation compared to none in the prorestriction category. Overall, there seemed to be very little interest in funding campaigns seeking to take political or legal action against mandates, particularly compared to the relative success seen in campaigns seeking medical care. An interesting subcategory consisted of the 7 campaigns in opposition to COVID-19 restrictions for health care workers, none of which achieved success, having raised very minimal funding despite asking for sizable donations. Among the campaigns written by or for those in the health care industry, all but one contained false or misleading claims about COVID-19. Health care workers continue to represent a challenging demographic, with significant regional variation in vaccination rates and opinions on COVID-19 treatments [43,44]. The lack of support and the small minority of health care workers seeking assistance may suggest increased COVID-19 awareness, limited appetite to fund such endeavors, or both.

The overall success rate of COVID-19–related campaigns approached 14%, indicating a sizeable and persistent gap between the funds raised and the funding needs of individuals still affected by COVID-19 [29]. Compared to a similar analysis by Saleh et al [2], in the first months of the pandemic (March 2020), this study shows a similar rate of success as measured by campaigns reaching their goal. In contrast to that study, we found a much higher median goal (US $10,000 vs US $5000) with a higher median amount raised (US $2,808 vs US $930). Two years into the pandemic, there is still support for campaigns at approximately the same level previously noted in the literature. Despite this continued support, there were substantial differences in the types of campaigns that received funding.

Limitations

Our study has several limitations. A major limitation is the singular time frame when the campaigns were collected. The goal was to obtain a snapshot of the landscape of COVID-19 crowdfunding, but only active or recently closed campaigns were captured. As the date range (2011-2021) suggests, some of these campaigns had been active for a long time while others were recently open. The older campaigns (including those that preexisted COVID-19) were subject to the same screening for relevance. We intentionally excluded campaigns open for <30 days to minimize this limitation. Conversely, it is possible that some of the most successful campaigns open and close quickly after meeting funding goals. This was not demonstrated in the sensitivity analysis when we reviewed campaigns open less than 30 days. Moreover, the description of "alternative treatments" has been presented in a singular study, and we take our definition from that study [45]. We sought to capture the broad interpretation of medical treatments that have not been adopted as standard of care in the treatment of COVID-19. Data are always evolving, and some or all of these treatments may become part of that standard in the future. Despite these limitations, this is the first cross-sectional analysis of COVID-19 misinformation among GoFundMe campaigns and the first study to examine the crowdfunding for political and legal challenges to COVID-19 mandates. Furthermore, it adds to the growing body of literature on social networking and crowdfunding for alternative treatments to COVID-19.

Conclusions

There was a small minority of crowdfunding campaigns seeking assistance for unproven alternative COVID-19 treatments, about half of which contained COVID-19 misinformation and very few of which were successful. Additionally, the majority of campaigns seeking funding for legal or political action against COVID-19 mandates contained COVID-19 misinformation and were unpopular and underfunded. Crucially, these data also raise concerns for the moral and ethical implications of the funds raised; where these monies used to fund the purposes stated or where they meant to capitalize on often politically charged beliefs? Our findings have implications for how individuals and organizations seek and obtain funding and add to the growing commentary on the ethical challenges associated with social media, COVID-19, and misinformation.

Acknowledgments

We thank Hartley Brody for assistance in obtaining GoFundMe data.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Key word categorization and search terms for each category.
[DOCX File , 21 KB - jmir_v24i7e38395_app1.docx]

Multimedia Appendix 2
Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist.
[PDF File (Adobe PDF File), 120 KB - jmir_v24i7e38395_app2.pdf]

References

1. Covid-19's $24 trillion cost (so far) means economics will never be the same. Forbes. 2021 Feb 26. URL: https://www.forbes.com/sites/williampesek/2021/02/26/covid-19s-24-trillion-cost-so-far-means-economics-will-never-be-the-same/?sh=771257bc4844 [accessed 2021-12-27]
2. Saleh SN, Lehmann CU, Medford RJ. Early Crowdfunding Response to the COVID-19 Pandemic: Cross-sectional Study. J Med Internet Res 2021 Feb 09;23(2):e25429 [FREE Full text] [doi: 10.2196/25429] [Medline: 33523826]
3. Rajwa P, Hopen P, Mu L, Paradysz A, Wojnarowicz J, Gross CP, et al. Online Crowdfunding Response to Coronavirus Disease 2019. J Gen Intern Med 2020 Aug 27;35(8):2482-2484 [FREE Full text] [doi: 10.1007/s11606-020-05896-x] [Medline: 32462566]
4. GoFundMe CEO: Hello Congress, Americans need help and we can't do your job for you. USA Today. URL: https://www.usatoday.com/story/opinion/voices/2021/02/11/gofundme-ceo-congress-pass-covid-relief-desperate-americans-column/4440425001 [accessed 2021-12-27]
5. Saleh SN, Ajufo E, Lehmann CU, Medford RJ. A Comparison of Online Medical Crowdfunding in Canada, the UK, and the US. JAMA Netw Open 2020 Oct 01;3(10):e2021684 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.21684] [Medline: 33104206]
6. Berliner LS, Kenworthy NJ. Producing a worthy illness: Personal crowdfunding amidst financial crisis. Soc Sci Med 2017 Aug;187:233-242. [doi: 10.1016/j.socscimed.2017.02.008] [Medline: 28274601]
7. Bassani G, Marinelli N, Vismara S. Crowdfunding in healthcare. J Technol Transf 2018 Apr 20;44(4):1290-1310. [doi: 10.1007/s10961-018-9663-7]
8. Snyder J, Mathers A, Crooks VA. Fund my treatment!: A call for ethics-focused social science research into the use of crowdfunding for medical care. Soc Sci Med 2016 Nov;169:27-30. [doi: 10.1016/j.socscimed.2016.09.024] [Medline: 27665200]
9. Gisondi MA, Barber R, Faust JS, Raja A, Strehlow MC, Westafer LM, et al. A Deadly Infodemic: Social Media and the Power of COVID-19 Misinformation. J Med Internet Res 2022 Feb 01;24(2):e35552 [FREE Full text] [doi: 10.2196/35552] [Medline: 35007204]
10. Karkowsky CE. Doctors Feel Compassion Fatigue About Vaccine Refusers. The Atlantic. 2021 Aug 11. URL: https://www.theatlantic.com/ideas/archive/2021/08/health-care-workers-compassion-fatigue-vaccine-refusers/619716 [accessed 2021-12-27]
11. Alharbi J, Jackson D, Usher K. The potential for COVID-19 to contribute to compassion fatigue in critical care nurses. J Clin Nurs 2020 Aug 18;29(15-16):2762-2764 [FREE Full text] [doi: 10.1111/jocn.15314] [Medline: 32344460]
12. Naeem SB, Bhatti R. The Covid-19 'infodemic': a new front for information professionals. Health Info Libr J 2020 Sep 13;37(3):233-239 [FREE Full text] [doi: 10.1111/hir.12311] [Medline: 32533803]
13. Muric G, Wu Y, Ferrara E. COVID-19 Vaccine Hesitancy on Social Media: Building a Public Twitter Data Set of Antivaccine Content, Vaccine Misinformation, and Conspiracies. JMIR Public Health Surveill 2021 Nov 17;7(11):e30642 [FREE Full text] [doi: 10.2196/30642] [Medline: 34653016]
14. Cruickshank I, Ginossar T, Sulskis J, Zheleva E, Berger-Wolf T. Content and Dynamics of Websites Shared Over Vaccine-Related Tweets in COVID-19 Conversations: Computational Analysis. J Med Internet Res 2021 Dec 03;23(12):e29127 [FREE Full text] [doi: 10.2196/29127] [Medline: 34665760]
15. 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]
16. Naeem SB, Bhatti R, Khan A. An exploration of how fake news is taking over social media and putting public health at risk. Health Info Libr J 2021 Jun 12;38(2):143-149 [FREE Full text] [doi: 10.1111/hir.12320] [Medline: 32657000]

17. Kłak A, Grygielska J, Mańczak M, Ejchman-Pac E, Owoc J, Religioni U, et al. Online Information of COVID-19: Visibility and Characterization of Highest Positioned Websites by Google between March and April 2020-A Cross-Country Analysis. Int J Environ Res Public Health 2022 Jan 28;19(3):1491 [FREE Full text] [doi: 10.3390/ijerph19031491] [Medline: 35162513]
18. Ahmed W, Vidal-Alaball J, Vilaseca J. A Social Network Analysis of Twitter Data Related to Blood Clots and Vaccines. Int J Environ Res Public Health 2022 Apr 11;19(8):4584 [FREE Full text] [doi: 10.3390/ijerph19084584] [Medline: 35457451]
19. Snyder J, Zenone M, Caulfield T. Crowdfunding Campaigns and COVID-19 Misinformation. Am J Public Health 2021 Apr;111(4):739-742. [doi: 10.2105/ajph.2020.306121]
20. At least 4 groups have been using PayPal and GoFundMe to fundraise for ivermectin, Covid misinformation. Forbes. 2021 Oct 07. URL: https://www.forbes.com/sites/abrambrown/2021/10/07/at-least-4-groups-have-been-using-paypal-and-gofundme-to-fundraise-for-ivermectin-covid-misinformation/?sh=1feab5e93ee3 [accessed 2022-03-24]
21. GoFundMe takes down $180,000 fundraiser for anti-mandate lawsuit, saying it violates their policy against vaccine 'misinformation'. Business Insider. 2021 Oct 11. URL: https://www.businessinsider.com/gofundme-takes-down-fundraiser-for-anti-vaccine-mandate-lawsuit-2021-10 [accessed 2022-03-24]
22. The lucrative business of stoking vaccine skepticism. Slate. URL: https://slate.com/technology/2021/08/givesendgo-gofundme-vaccine-skeptics-covid-misinformation.html [accessed 2022-03-27]
23. Covid-19 — Myth Versus Fact. John Hopkins Medicine. 2021. URL: https://www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus/2019-novel-coronavirus-myth-versus-fact [accessed 2022-03-23]
24. Myths and Facts about COVID-19 Vaccines. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/vaccines/facts.html [accessed 2022-03-23]
25. Coronavirus disease (COVID-19) advice for the public: Mythbusters. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters [accessed 2022-03-23]
26. List of unproven methods against COVID-19. Wikipedia. URL: https://en.wikipedia.org/wiki/List_of_unproven_methods_against_COVID-19 [accessed 2022-03-23]
27. COVID-19 myths debunked. Mayo Clinic Health System. 2020 Nov 03. URL: https://www.mayoclinichealthsystem.org/hometown-health/featured-topic/11-covid-19-myths-debunked [accessed 2022-03-23]
28. Cohen AJ, Brody H, Patino G, Ndoye M, Liaw A, Butler C, et al. Use of an Online Crowdfunding Platform for Unmet Financial Obligations in Cancer Care. JAMA Intern Med 2019 Dec 01;179(12):1717. [doi: 10.1001/jamainternmed.2019.3330]
29. Igra M, Kenworthy N, Luchsinger C, Jung J. Crowdfunding as a response to COVID-19: Increasing inequities at a time of crisis. Soc Sci Med 2021 Aug;282:114105 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114105] [Medline: 34139479]
30. Understanding the infodemic and misinformation in the fight against COVID-19. Pan American Health Organization. URL: https://iris.paho.org/bitstream/handle/10665.2/52052/Factsheet-infodemic_eng.pdf [accessed 2022-07-14]
31. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE). Epidemiology 2007;18(6):805-835. [doi: 10.1097/ede.0b013e3181577511]
32. COVID-19 and “alternative” treatments: what you need to know. National Center for Complimentary and Integrative Health. URL: https://www.nccih.nih.gov/health/covid-19-and-alternative-treatments-what-you-need-to-know [accessed 2022-03-28]
33. Andrejko KL, Pry JM, Myers JF, Fukui N, DeGuzman JL, Openshaw J, California COVID-19 Case-Control Study Team. Effectiveness of Face Mask or Respirator Use in Indoor Public Settings for Prevention of SARS-CoV-2 Infection - California, February-December 2021. MMWR Morb Mortal Wkly Rep 2022 Feb 11;71(6):212-216 [FREE Full text] [doi: 10.15585/mmwr.mm7106e1] [Medline: 35143470]
34. Preston LE, Chevinsky JR, Kompaniyets L, Lavery AM, Kimball A, Boehmer TK, et al. Characteristics and Disease Severity of US Children and Adolescents Diagnosed With COVID-19. JAMA Netw Open 2021 Apr 01;4(4):e215298 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.5298] [Medline: 33835179]
35. Clinical spectrum of SARS-CoV-2 infection. National Institutes of Health. 2021 Oct 19. URL: https://www.covid19treatmentguidelines.nih.gov/overview/clinical-spectrum [accessed 2022-03-30]
36. National COVID- 19 preparedness plan. The White House. URL: https://www.whitehouse.gov/covidplan [accessed 2022-03-30]
37. Anand P, Stahel VP. Review the safety of Covid-19 mRNA vaccines: a review. Patient Saf Surg 2021 May 01;15(1):20 [FREE Full text] [doi: 10.1186/s13037-021-00291-9] [Medline: 33933145]
38. Al Khames Aga QA, Alkhaffaf WH, Hatem TH, Nassir KF, Batineh Y, Dahham AT, et al. Safety of COVID-19 vaccines. J Med Virol 2021 Dec 28;93(12):6588-6594 [FREE Full text] [doi: 10.1002/jmv.27214] [Medline: 34270094]
39. Pertwee E, Simas C, Larson HJ. An epidemic of uncertainty: rumors, conspiracy theories and vaccine hesitancy. Nat Med 2022 Mar;28(3):456-459. [doi: 10.1038/s41591-022-01728-z] [Medline: 35273403]
40. Yeung AWK, Tosevska A, Klager E, Eibensteiner F, Tsagkaris C, Parvanov ED, et al. Medical and Health-Related Misinformation on Social Media: Bibliometric Study of the Scientific Literature. J Med Internet Res 2022 Jan 25;24(1):e28152 [FREE Full text] [doi: 10.2196/28152] [Medline: 34951864]

41. Ghaddar A, Khandaqji S, Awad Z, Kansoun R. Conspiracy beliefs and vaccination intent for COVID-19 in an infodemic. PLoS One 2022;17(1):e0261559 [FREE Full text] [doi: 10.1371/journal.pone.0261559] [Medline: 35020721]
42. Gisondi M, Chambers D, La TM, Ryan A, Shankar A, Xue A, et al. A Stanford Conference on Social Media, Ethics, and COVID-19 Misinformation (INFODEMIC): Qualitative Thematic Analysis. J Med Internet Res 2022 Feb 15;24(2):e35707 [FREE Full text] [doi: 10.2196/35707] [Medline: 35030089]
43. Sharma R, Kumar M, Rohilla KK. COVID-19 Infodemic: Evaluating Information-Seeking Behaviour Among Healthcare Workers During a Pandemic. Cureus 2022 Jan;14(1):e20910 [FREE Full text] [doi: 10.7759/cureus.20910] [Medline: 35154911]
44. Reses HE, Jones ES, Richardson DB, Cate KM, Walker DW, Shapiro CN. COVID-19 vaccination coverage among hospital-based healthcare personnel reported through the Department of Health and Human Services Unified Hospital Data Surveillance System, United States, January 20, 2021-September 15, 2021. Am J Infect Control 2021 Dec;49(12):1554-1557 [FREE Full text] [doi: 10.1016/j.ajic.2021.10.008] [Medline: 34802705]
45. Song S, Cohen AJ, Lui H, Mmonu NA, Brody H, Patino G, et al. Use of GoFundMe to crowdfund complementary and alternative medicine treatments for cancer. J Cancer Res Clin Oncol 2020 Jul 26;146(7):1857-1865. [doi: 10.1007/s00432-020-03191-0] [Medline: 32219517]

Abbreviations

IRB: Institutional Review Board
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by M Gisondi, J Faust; submitted 31.03.22; peer-reviewed by N Hu; comments to author 30.05.22; revised version received 03.06.22; accepted 04.06.22; published 27.07.22.

Please cite as:
Shaw NM, Hakam N, Lui J, Abbasi B, Sudhakar A, Leapman MS, Breyer BN
COVID-19 Misinformation and Social Network Crowdfunding: Cross-sectional Study of Alternative Treatments and Antivaccine Mandates
J Med Internet Res 2022;24(7):e38395
URL: https://www.jmir.org/2022/7/e38395
doi:10.2196/38395
PMID:35820053

Research Letter

Using Machine Learning to Efficiently Vaccinate Homebound Patients Against COVID-19: A Real-time Immunization Campaign

Anish Kumar[1*], BA; Jennifer Ren[1*], BA; Katherine A Ornstein[2,3], MPH, PhD; Peter M Gliatto[4], MD

[1]Icahn School of Medicine at Mount Sinai, New York, NY, United States

[2]Department of Geriatrics and Palliative Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[3]Institute for Care Innovations at Home, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[4]Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[*]these authors contributed equally

Corresponding Author:
Anish Kumar, BA
Icahn School of Medicine at Mount Sinai
1 Gustave L. Levy Place
New York, NY, 10029
United States
Phone: 1 9085668399
Email: anish.kumar@icahn.mssm.edu

(J Med Internet Res 2022;24(7):e37744) doi:10.2196/37744

KEYWORDS

home care; covid; vaccination; COVID-19; machine learning; vaccine; geographic cluster; patient data; clustering algorithm; geospatial; digital surveillance; public health; logistic operation; electronic health record integration

Introduction

In March 2021, the New York City Department of Health and Mental Hygiene (NYCDOHMH) received a supply of single-dose Janssen COVID-19 vaccines to vaccinate the city's patients who are homebound, but they needed assistance to identify and reach this population [1,2]. Mount Sinai Visiting Doctors (MSVD) is a large home-based primary care program serving more than 1200 patients who are homebound throughout Manhattan. We partnered with NYCDOHMH to vaccinate patients in our program. The administrative team generally schedules routine home visits based on zip codes, using 12 unique catchment areas covering all of Manhattan. This existing zoning system was inadequate for vaccination purposes for several reasons, mainly because these zones were too large. Furthermore, the additional task of manually scheduling patients by zone would be overwhelming for administrative staff, especially given the temporal constraints of the Janssen vaccine (ie, doses expired in 6 hours). In response, we developed a system to geographically cluster patients to efficiently vaccinate our homebound patients against COVID-19.

Methods

Ethics Approval

The Icahn School of Medicine at Mount Sinai's Program for the Protection of Human Subjects approved and granted a waiver of consent for this secondary data analysis study (STUDY-21-00157) which was conducted in accordance with the Helsinki Declaration.

Overview

We developed a software program that takes a cohort of unvaccinated patients in the MSVD Program as input and assigns each patient a cluster number. We used Python 3 (Python Software Foundation) [3] to process patient addresses, group them into clusters, and export these clusters into a standard database format and onto a map. Specifically, we used an open-source implementation of a modified unsupervised K-means clustering machine learning algorithm to group patients [4].

The first step was to convert patient households to spatial representations. Using the Google Maps Geocoding application programming interface (API), we obtained latitude/longitude coordinates of patient residences, which served as the input data to our algorithm.

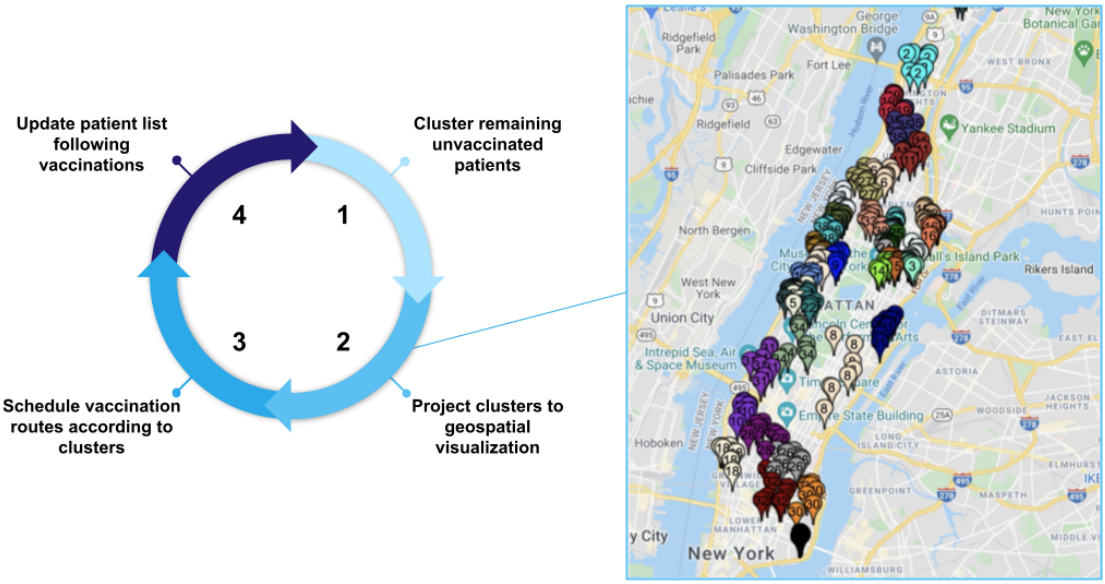
These coordinates were then fed to the modified unsupervised K-means clustering algorithm. The standard K-means clustering algorithm clusters data points into *K* groups, while the modified algorithm allows one to impose constraints on cluster size, ensuring each group contains a number of data points within a specified range [4]. This was important for our process to avoid assigning an excessive number of patients to any one cluster; although many patients live close to each other, provider routes could not exceed 6 hours. Targeting approximately 5 patients

per provider and bearing in mind that not all identified patients would desire and be available when the vaccine was offered by the scheduling team, we enforced a minimum of 12 and a maximum of 15 patients per cluster.

Next, we projected the data points onto a geospatial visualization for easy interpretability by the scheduling team and providers. Using the Google Maps API, we generated an HTML page that contained pins with distinct numbers and colors representing the cluster to which a patient belonged as well as the patient's name.

Our process was iterative: every week after patients were scheduled and immunized, we updated our census to reflect the remaining number of patients who were unvaccinated. The team called patients to confirm appointments and updated the patient list. We then re-ran our clustering process, integrating any changes to cluster size preferences, for the next round of immunizations for patients who are homebound (Figure 1).

Figure 1. An example of a vaccination campaign, using mock patient data, for patients who are homebound in the Mount Sinai Visiting Doctors Program routing process. It was iterative week-to-week and included the following steps: (1) patients who were unvaccinated were identified and clustered into groups, (2) patient addresses were pinned on a map with identifying group labels, (3) patients were called for vaccine scheduling according to group number, and (4) the patient list was updated to reflect who remained unvaccinated.



Results

Table 1 summarizes the demographic information of the patients and the results of our campaign between March and April 2021, which involved about 100 vaccination clusters and routes over 6 weeks. On average, we vaccinated 5.6 patients per provider per day, averaging 22.1 total patients per day [5]. Each provider accomplished their assigned route within the time constraints of the Janssen vaccine, and no doses were wasted.

Table 1. Demographics and statistics from our COVID-19 homebound vaccination campaign of Manhattan-based patients in the Mount Sinai Visiting Doctors Program.

	Value
Patient demographics	
Patients who are homebound	428
Average age (years)	83.9
Average Elixhauser comorbidity score [6][a]	3.8
Sex, n (%)	
Female	323 (75.5)
Male	105 (24.5)
Racial/ethnic identity, n (%)	
White	148 (34.6)
Black or African American	63 (14.7)
Asian	15 (3.5)
Hispanic	108 (25.2)
Other	86 (20.1)
Unknown	8 (1.9)
Patient's family members and caregivers, n	92
Vaccination campaign statistics	
Providers vaccinating per day (n), range	3-6
Average number of patients vaccinated per day	22.1
Average duration of provider time spent vaccinating (hours)	4.6
Average duration of individual vaccination (including transit time, vaccine administration, and 15-minute postvaccination observation time; minutes)	52

[a]Elixhauser scores were available for 372 of the patients who are homebound.

Discussion

The new tool optimized logistic operations for an acute public health intervention while minimizing wasted resources. The model was later used to quickly deploy booster immunizations during the surge of the Omicron subvariant from December 2021 to January 2022. Future research will consider the ability to create routes with ordered stops given a provider's choice of transportation and more flexibility in the cluster size depending on the population density of a given region. This feature is especially important in densely populated cities, where providers are unlikely to travel by car. Additional future work can include electronic health record integration for more streamlined access and allowing scheduling teams to directly recluster patients.

Ultimately, the newly developed approach was instrumental in maximizing efficiency and minimizing vaccine waste, suggesting its potential for future use in home-based health care or other public health interventions. The success of this project further demonstrates the value of novel technological approaches in improving the efficiency of clinical operations.

Conflicts of Interest

None declared.

References

1. Otterman S. Homebound and waiting: vaccine push misses some older New Yorkers. The New York Times. 2021 Apr 23. URL: https://www.nytimes.com/2021/04/23/nyregion/homebound-vaccine-nyc.html [accessed 2021-11-18]
2. Baulkman J. Vaccinating homebound patients is an uphill battle. MDedge Family Medicine. 2021 Apr 23. URL: https://www.mdedge.com/familymedicine/article/239068/coronavirus-updates/vaccinating-homebound-patients-uphill-battle [accessed 2021-11-18]
3. Van Rossum G, Drake FL. Python 3 Reference Manual. Scotts Valley, CA: CreateSpace; 2009.
4. Bradley PS, Bennet KP, Demiriz A. Constrained k-means clustering. Microsoft Res 2000:1-8.

5. Gliatto P, Franzosa E, Chavez S, Ng A, Kumar A, Ren J, et al. Covid-19 vaccines for homebound patients and their caregivers. New Engl J Med Catalyst 2021:14.
6. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Med Care 1998 Jan;36(1):8-27. [doi: 10.1097/00005650-199801000-00004] [Medline: 9431328]

Abbreviations

API: application programming interface
MSVD: Mount Sinai Visiting Doctors
NYCDOHMH: New York City Department of Health and Mental Hygiene

Edited by T Leung; submitted 04.03.22; peer-reviewed by O Rahaman, FM Calisto; comments to author 11.05.22; revised version received 04.06.22; accepted 07.06.22; published 12.07.22.

Please cite as:
Kumar A, Ren J, Ornstein KA, Gliatto PM
Using Machine Learning to Efficiently Vaccinate Homebound Patients Against COVID-19: A Real-time Immunization Campaign
J Med Internet Res 2022;24(7):e37744
URL: https://www.jmir.org/2022/7/e37744
doi:10.2196/37744
PMID:35679053

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org